(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 April 2003 (03.04.2003)

PCT

(10) International Publication Number
WO 03/027682 A2

(51) International Patent Classification[7]: G01N 33/68, 33/58

(21) International Application Number: PCT/US02/30742

(22) International Filing Date:
27 September 2002 (27.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/325,335   27 September 2001 (27.09.2001)   US

(71) Applicant: PURDUE RESEARCH FOUNDATION [US/US]; 1291 Cumberland Avenue, West Lafayette, IN 47906 (US).

(71) Applicants and
(72) Inventors: REGNIER, Fred, E. [US/US]; 1219 Tuckahoe Lane, West Lafayette, IN 49706 (US). ZHANG, Roujian, [CN/US]; 587 N. Ventu Park Road, Ste E PMB 126, Newberry Park, CA 91320 (US). CHAKRABORTY, Asish, B. [BD/US]; 1 Pheasant Circle, Apt. 11, Milford, MA 01757 (US).

(74) Agent: SANDBERG, Victoria, A.; Mueting, Raasch & Gebhardt, P.A., P.O. Box 581415, Minneapolis, MN 55458-1415 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MATERIALS AND METHODS FOR CONTROLLING ISOTOPE EFFECTS DURING FRACTIONATION OF ANALYTES

(57) Abstract: Compositions and methods for controlling or eliminating isotope effects during fractionation of chemically equivalent but isotopically distinct compounds. Isotope coding agents contain heavy isotopes other than deuterium. The invention facilitates intelligent data acquisition. After sample fractionation, isotope abundance ratios are calculated using mass spectrometry, and analytes of interest are identified in real time.

WO 03/027682 A2

MATERIALS AND METHODS FOR CONTROLLING ISOTOPE EFFECTS DURING FRACTIONATION OF ANALYTES

5   This application claims the benefit of U.S. Provisional Application Serial No. 60/325,335, filed September 27, 2001, which is incorporated herein by reference in its entirety.

STATEMENT OF GOVERNMENT RIGHTS

This invention was made with government support under a grant from
10   the National Institutes of Health, Grant No. GM 59996. The U.S. Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

The science of proteomics seeks to efficiently quantify, identify and
15   characterize the large number of proteins/peptides that characterize a biological system. Two approaches to proteomics are being pursued today. One focuses on cataloging all the proteins in a biological system, the cellular components with which each of these protein interact, the pathways of which they are a part, and the location in which they reside. The other strategy, known as
20   "comparative proteomics" is based on a comparison of proteins in a biological system in two different states. Disease and a wide range of other stimuli cause biological systems to pass into a new, chemically distinct state distinguished by changes in the occurrence and amount of specific proteins. Comparing samples taken from organisms in the normal and an altered state can be used to
25   recognize proteins involved in the transition.

Although "proteomics" is a recently coined term, comparative proteomics has actually been practiced for at least three decades. For example, early studies used 2-D gel electrophoresis to examine thousands of proteins in blood. These reports, and many since, depended on resolution of complex
30   protein mixtures with 2-D gels and differential staining to compare samples and recognize differences. This classical method is labor intensive, quantification is poor, and it is difficult to identify spots thought to be important.

With the advent of huge DNA and protein sequence databases there is an increasing dependence on mass spectrometry to identify proteins and peptides obtained from separation systems. The speed and resolution of mass spectrometers is shifting the focus in proteomics toward more rapid delivery of

5   peptide mixtures to mass spectrometers and obtaining quantitative as well as qualitative data during mass analysis.

Stable isotope coding strategies are of great value for distinguishing changes in comparative proteomics. These coding techniques may be broadly characterized as internal standard methods in which components from control

10   samples are derivatized with an isotopically distinct coding agent, mixed with experimental samples, then used as standards for determining the relative concentration of components in experimental samples derivatized with a different isoform of the coding agent (PCT WO 01/86306, published November 15, 2001, Ji et al., *J. Chromatogr. B*, (2000) 745, 197-210). Most of the coding

15   agents used today are labeled with deuterium, and relative concentration measurements are based on isotope abundance ratio determinations with either matrix assisted laser desorption ionization-mass spectrometry (MALDI-MS) or electrospray ionization-mass spectrometry (ESI-MS).

As with all internal standard methods, it is important that the behavior of

20   analytes and standards be as nearly alike as possible before the final step of abundance ratio measurement. Ideally, segregation would occur only in the final step during quantification. The attractive feature of creating internal standards through isotopic labeling is that discrimination is minimized, particularly when the internal standard and analyte vary by a single heavy atom.

25   The problem with current stable isotope coding methods for proteomics is that as the number of deuterium atoms is increased to enlarge the mass difference between isotopically coded standards and analytes, there is a corresponding increase in chromatographic resolution of the isotopic isoforms, particularly in the case of reversed-phase chromatography (Zhang et al., Anal.

30   Chem., 2001, 73, 5142-5149). As a result, the concentration ratio of isoforms varies continuously across the elution profile of the two components.

This "isotope effect" has a number of undesirable consequences. Because of the isotope effect caused by deuterium in chromatographic

2

separation, the accuracy of abundance ratio measurement is greatly compromised, particularly in MALDI-MS. In online ESI-MS analysis, there is a serious trade-off between accurate quantification and MS/MS peptide sequencing because of the deuterium isotope effect.

5   For example, using existing methods (e.g., abundance ratio measurements) to determine a relative change in concentration of components from a single mass spectrum is not possible. Instead, relative concentration must be obtained by a comparison of area measurements between integrated extracted ion chromatograms in the case of ESI-MS or from eluate fractions

10   with MALDI-MS. High quality MS/MS data of peptides are crucial to the reliability of protein identification and characterization. Unfortunately, this requires certain tuning and ion accumulation. In another word, it takes time. It is very much desired to selectively fragment the peptides of interest.

       In the case of ESI-MS, the analyte peak must completely elute from the

15   LC-MS system before it can be determined whether the peptide has changed in concentration and is therefore of interest (Griffin et al., Journal of the American Society for Mass Spectrometry, 2001, 12, 1238-1246). This effectively precludes the use of on-line, intelligent data acquisition and analysis (IDA), also referred to as real-time data dependent analysis (DDA). Either a second

20   chromatographic run is required after the abundance ratio is determined in order to selectively perform MS/MS analysis on the peptides of interest, or the mass spectrometer has to acquire MS/MS data on every peptide to be sure data has been acquired on peptides that changed in concentration. In either case, instrument and computer time are wasted.

25   Feedback control software is available in some commercially available mass spectrometers. The most abundant peaks are typically selected for further analysis on-line in "real time". The most abundant peaks, however, do not necessarily coincide with analytes that are significantly up or down regulated, which in comparative proteomics are more often the analytes of interest. In

30   other commercially available instruments, LC/MS is run twice, first to obtain abundance ratios and second to selectively fragment peptides based on the abundance ratios. This method may yield information on analytes that are up or down regulated, but has a major drawback in that abundance ratio based

feedback control is not performed in "real time" since two chromatographic runs are used.

Reconstruction of extracted ion chromatograms is really only possible with ESI-MS. Reconstruction of peaks from MALDI-MS is very difficult in
5   complex mixtures unless 10-60 fractions are collected across each peak. Anywhere from 2,000-10,000 fractions would have to be collected and analyzed by MALDI-MS to reconstruct extracted ion chromatograms for the thousands of peptides encountered in a single reversed phase chromatographic separation. This is so cumbersome that MALDI-MS by any approach but continuous
10 deposition is essentially precluded when quantification accuracy is an issue.

Integration of peak areas is even more difficult and inaccurate when isotopically labeled peptides are fractionated in one or more of the early steps in a multi-dimensional separation experiment. For example, isotopically labeled peptides could be separated in ion exchange chromatography followed by
15 reversed-phase chromatography, or reversed-phase chromatography followed by ion mobility separation.

Further complications can arise when ionization efficiency of the isoforms varies with time in ESI-MS or between fractions in MALDI-MS. Suppression of ionization between peptides has been noted in ESI-MS when
20 total peptide concentration is high, as when one peptide is eluting in a large background of another. This means that ionization efficiency can vary across a chromatographic peak. In MALDI-MS, the chromatographic fractions used for spectral analysis may be enriched in one isotopic isoform over the other and may differ widely in matrix components. The peptide isoforms could be
25 suppressed to very different degrees when they are not eluted simultaneously and thus ionized with a different matrix (i.e., with other co-eluting peptides). These effects compromise abundance ratio measurements by potentially producing both significant systematic errors and a higher level of random errors (Zhang et al., Anal. Chem. 2001, 73, 5142 –5149), even when relative peak
30 areas are estimated with extracted ion chromatograms.

Finally there is the most serious problem of all. When the isotopic isoforms are completely resolved, as is frequently the case with labeled peptides that contain many deuterium atoms, whether or not the two peaks are related

cannot be determine unless the peptides are sequenced or mass is measured with very high mass accuracy. This leads to the erroneous conclusion that the singlet cluster seen in one of the resolved peaks is representative for the peptide and that it has undergone a major change in concentration.

5   Clearly, minimizing chromatographic resolution of analyte isoforms would improve measurement accuracy and enable real-time measurements.

10   SUMMARY OF THE INVENTION

The invention advances the science of proteomics as well as analytical chemistry in general in that it provides a reliable, real-time method to detect the changes in analyte concentration and/or modification. Improved accuracy of isotope abundance ratio measurement is achieved, and false positives and false

15   negatives caused by the systematic errors in the measurement are reduced or eliminated.

Isotope effect can cause significant errors in protein/peptide quantitation and identification. The invention solves this problem by minimizing the isotope effect. The isotope effect can be reduced by eliminating deuterium (^{2}H) from

20   the isotope coding agent and using instead different heavy isotopes, preferably ^{13}C, ^{18}O and/or ^{15}N, more preferably ^{13}C, to provide the difference in mass. Preferably, proteins/peptides are labeled with an isotope coding agent containing two or more, preferably three or more, ^{13}C atoms, ^{18}O atoms, ^{15}N atoms or any combination thereof. A difference of at least 2-3 amu (atomic

25   mass units) among isoforms is needed, preferably a difference of 3 or more amu, otherwise the isotopically labeled analytes may heavily overlap in mass spectra and become difficult to deconvolute. The isotope coding agent of the invention preferably contains no deuterium atoms.

Importantly, the present invention makes real-time intelligent data

30   acquisition possible. In addition, the identical or nearly identical chromatographic retention times can be used to confirm the recognition of isoform pairs. The invention is well-suited to comparative proteomics

applications, general analytical chemistry, and the analysis of biomolecules such as peptides, small molecules, sugars and drugs.

In one aspect, the invention provides an isotope coding agent that includes at least three non-deuterium isotopes. The non-deuterium isotopes present in the isotope coding agent are preferably ^{13}C, ^{18}O, ^{15}N or any combination thereof. The non-deuterium isotopes that are incorporated into an isotope coding agent of the invention can differ or be the same. A preferred isotope coding agent contains a reactive functional group that reacts with the analyte of interest, preferably with an amine, a carboxyl, a hydroxyl or a thiol present on the analyte of interest. Examples of a reactive functional group are iodoacetamide and iodoacetate, which react with free thiols. In a particularly preferred embodiment the isotope coding agent includes a reactive functional group and an isotopic linker that includes the least three heavy non-deuterium isotopes. Optionally, the isotope coding agent further includes an affinity "tag", i.e., a functional group such as biotin for affinity selection. An example of such an isotope coding agent is the ^{13}C-ICAT reagent available from Applied Biosystems, Inc. Succinic anhydride, N-acetoxysuccinimide and propionate-N-hydroxysuccinimide that include heavy non-deuterium isotopes are also preferred isotope coding agents.

The invention also includes a method for making an isotope coding agent. Many deuterated isotope coding agents are in common use. In this aspect of the invention, any deuterated isotope coding agent can serve as the basis for the design of an isotope coding agent that includes non-deuterium heavy isotopes. The method involves synthesizing an isoform of the isotope coding agent that includes non-deuterium heavy isotopes instead of deuterium. At least three heavy isotopes are independently selected from the group consisting of ^{13}C, ^{18}O and ^{15}N and incorporated into a compound to yield the analogous ^{13}C-, ^{18}O- and/or ^{15}N-containing isotope coding agent. The isotope effect observed for the ^{13}C-, ^{18}O- and/or ^{15}N- isotope coding agent is expected to be smaller than the isotope effect observed for the deuterated isoform. Also included in the invention is a method for isotopically coding an analyte that involves covalently linking the analyte to an isotope coding agent, wherein the

isotope coding agent includes at least three heavy isotopes independently selected from the group consisting of ^{13}C, ^{18}O and ^{15}N.

In another aspect, the invention includes a peptide that is covalently linked to one or more isotope coding agents, each having at least three non-

5   deuterium (e.g., ^{13}C, ^{18}O or ^{15}N or any combination thereof) isotopes. Preferred isotope coding agents for linkage to the peptide include an ICAT reagent, succinic anhydride, N-acetoxysuccinimide and propionate-N-hydroxysuccinimide.

In yet another aspect, the invention provides a device for detecting a

10   difference in the concentration of an analyte present in first sample and second samples. The device includes a sample fractionator which has an outlet, and a mass spectrometer coupled to the outlet of the sample fractionator. The fractionator is preferably a chromatography column, more preferably a reversed phase column.

15   Software is also included, which advantageously allows determination of the abundance ratio for the analyte using the mass spectrum of a combined sample immediately following elution of isotopically labeled analytes from the sample fractionator. In one embodiment, the software determines a normalized abundance ratio characterizing analytes whose concentration is the same in the

20   first and second samples and an abundance ratio of the first and second isotopically labeled analytes. A difference in the abundance ratio of the first and second isotopically labeled analytes and the normalized abundance ratio is indicative of a difference in concentration of the analyte in the first and second samples.

25   Importantly, the abundance ratios are determined in real time such that an eluted fraction containing analytes that have changed in concentration can be immediately subjected to further analysis using MS-MS. This analysis can be performed on-line, in real time. Advantageously, the decision of whether to perform MS-MS on a peak can be made by the software without input from a

30   human operator. Preferred devices are those incorporating software designed for proteomics applications and blood sample analysis.

A method for using the device is also provided. Samples to be analyzed each contain a plurality of analytes. Analytes in each sample are labeled with,

7

or contain, an isoform of an isotope coding agent. Preferably the isotope coding agent contains no deuterium. The isoforms of the isotope coding agent differ as between the samples; at least one isoform includes at least one non-deuterium heavy atom (e.g., ^{13}C, ^{18}O and/or ^{15}N) and has a mass of at least 3 amu greater
5   than the mass of an isoform comprising no heavy isotopes.

The samples are combined, and the combined sample is fractionated using the fractionator, yielding a plurality of elution peaks. Fractionation includes, for example, single or multi-dimensional reversed phase chromatography, ion exchange chromatography, hydrophobic interaction
10   chromatography, size exclusion chromatography, capillary gel electrophoresis, capillary zone electrophoresis, and capillary electrochromatography, capillary isoelectric focusing, immobilized metal affinity chromatography, affinity electrophoresis or any combination thereof. Preferably, the different isoforms of an analyte do not resolve during fractionation.
15   MS is performed on each elution peak, and the relative abundance ratio is calculated in real time to detect analytes exhibiting a change in concentration. Optionally, MS-MS is performed on the elution peak comprising the isoforms exhibiting a change in concentration to further analyze said isoforms. Preferably, mass spectrometric analysis is performed using matrix assisted laser
20   desorption ionization (MALDI), electrospray ionization (ESI), fast atom bombardment (FAB), electron impact ionization, atmospheric pressure chemical ionization (APCI), time-of-flight (TOF), quadrapole, ion trap, magnetic sector, ion cyclotron resonance mass, or combinations thereof. Also optionally, the analyte exhibiting the change in concentration is identified.
25   The first sample can be obtained from an environment before application of a stimulus, and the second sample can be obtained from an environment after application of the stimulus. Likewise, the first sample can be obtained from an environment in the presence of, or after application of, a stimulus and the second sample can be obtained from an environment in the absence of the
30   stimulus. The samples can be obtained from different organisms, cells, organs, tissues or bodily fluids, the method further comprising determining differences in concentration of at least one protein in the organisms, cells, organs, tissues or bodily fluids from which the samples were obtained, and can contain about 100

8

or more proteins. In biological samples, the analytes can be peptides produced by proteolysis of samples containing proteins, and MS-MS analysis of the elution peak that contains isoforms of peptides exhibiting a change in concentration can be used to identify the proteins from which the peptides were

5   derived.

The method of the invention allows detection of a difference in the concentration of an analyte, preferably a protein, present in first sample and second samples, wherein each sample includes a plurality of analytes. In this embodiment, the method involves first covalently attaching a first isoform of a

10   labeling agent to the analyte in the first sample to yield at least one first isotopically labeled analyte, and covalently attaching a second isoform of the labeling agent to the analyte in the second sample to yield at least one second isotopically labeled analyte, wherein the first isoform includes at least one non-deuterium heavy isotope, and wherein the first and second isoforms differ in

15   mass by at least 3 amu. At least portions of the first and second samples are then mixed to yield a combined sample.

The combined sample is subjected to mass spectrometric analysis to determine a normalized abundance ratio characterizing analytes whose concentration is the same in the first and second samples and an abundance ratio

20   of the first and second isotopically labeled analytes, wherein a difference in the abundance ratio of the first and second isotopically labeled analytes and the normalized abundance ratio is indicative of a difference in concentration of the analyte in the first and second samples.

Optionally, the method includes fractionating the combined sample

25   using a sample fractionator to yield at least one fraction containing the first and second isotopically labeled analytes prior to determining the abundance ratios. The isotope effect observed during fractionation of a mixture of the first and second isotopically labeled analytes is expected to be smaller than the isotope effect observed during fractionation of a mixture that includes the second

30   isotopically labeled analyte and an analyte labeled with a deuterated isoform of the labeling agent.

Optionally, mass spectrometric analysis is performed on the fraction immediately following elution of the fraction from the sample fractionator.

9

Abundance ratios are determined in real time such that an eluted fraction comprising analytes that have changed in concentration can be immediately subjected to further analysis using MS-MS.

When the analyte of interest is a protein, proteins in the samples can be

5   cleaved after labeling to yield first and second isotopically labeled peptides in the first and second samples, respectively. At least portions of the first and second samples are mixed to yield the combined sample, which is subjected to further fractionation and mass spectrometric analysis in accordance with the invention, as desired.

10   Alternatively, proteins in the samples can be cleaved prior to labeling to yield constituent peptides. A first isoform of a labeling agent is covalently attached to a peptide in the first sample to yield at least one first isotopically labeled peptide and a second isoform of the labeling agent is covalently attached to a peptide in the second sample to yield at least one second isotopically

15   labeled peptide. Preferably, the first and second isoforms of the labeling agent are attached to at least one amino group on peptides in the first and second samples. The first isoform comprises at least one non-deuterium heavy isotope, and the first and second isoforms differ in mass by at least 3 amu. At least portions of the first and second samples are then mixed to yield the combined

20   sample, which is subjected to further fractionation and mass spectrometric analysis in accordance with the invention, as desired.

A normalized abundance ratio characterizing peptides derived from proteins whose concentration is the same in the first and second samples is determined, as well as an abundance ratio of the first and second isotopically

25   labeled peptides. Mass spectrometric analysis is performed on the fraction immediately following elution of the fraction from the sample fractionator, and abundance ratios are determined in real time such that an eluted fraction comprising analytes that have changed in concentration can be immediately subjected to further analysis using MS-MS. A difference in the abundance ratio

30   of the first and second isotopically labeled peptides and the normalized abundance ratio is indicative of a difference in concentration in the first and second samples of a protein from which the peptide is derived.

10

Optionally, the first and second isotopically labeled peptides can include at least one affinity ligand. In that case, the method further includes, prior to determining the abundance ratios, contacting the peptides with a capture moiety to select peptides comprising the at least one affinity ligand. Also optionally,

5   the protein from which the peptide was derived can be identified.

In another aspect, the invention provides a multiplexed method for detecting a difference in the concentration of an analyte of interest present in multiple samples, each sample comprising a plurality of analytes. Preferably, about 3 to about 5 samples are analyzed concurrently. The method includes, for

10   each sample, covalently attaching an isoform of a labeling agent to the analyte of interest to yield an isotopically labeled analyte of interest. Different isoforms of the labeling agent are used for each sample, and the isoforms differ from one another in mass by at least about 3 amu. Preferably, all but at most one isoform include at least one non-deuterium heavy isotope. At least portions of the

15   samples are mixed to yield a combined sample.

The combined sample is subjected to mass spectrometric analysis to determine a normalized abundance ratio characterizing analytes whose concentration is the same in the samples and an abundance ratio of at least one pair of the isotopically labeled analytes of interest. A difference in the

20   abundance ratio of the isotopically labeled analytes of interest and the normalized abundance ratio is indicative of a difference in concentration of the analyte of interest in the samples.

Where the analyte of interest is a protein, the multiplexed method optionally includes cleaving the proteins in the samples to yield at least one

25   peptide of interest derived from the protein of interest. Then, for each sample, an isoform of a labeling agent is covalently attached to the peptide of interest to yield an isotopically labeled peptide of interest.

At least portions of the samples are mixed to yield a combined sample. The combined sample is subjected to mass spectrometric analysis to determine a

30   normalized abundance ratio characterizing peptides derived from proteins whose concentration is the same in the samples and an abundance ratio of at least one pair of the isotopically labeled peptides of interest. A difference in the abundance ratio of the isotopically labeled peptides of interest and the

11

normalized abundance ratio is indicative of a difference in concentration in the samples of a protein from which the peptide is derived. Again, preferably about 3 to 5 samples are analyzed concurrently.

In yet another aspect, the invention provides a method for identifying
5   isoforms of an analyte. Nondeuterated isoforms of the analyte do not resolve during fractionation, and the fractionation yields a plurality of elution peaks. MS can be performed on each elution peak to detect isoforms of the analyte.

BRIEF DESCRIPTION OF THE DRAWINGS

10

Figure 1 depicts a simulation based on Eqs. 4 and 7 of abundance ratio ($\blacklozenge$) as a function of elution time in a chromatographic separation of isotopically labeled peptides when $ratio_{true} = 1$. (a) when $R = 0$, $ratio_{obs}$ ($\blacklozenge$) does not vary with time and always equals $ratio_{truc}$; (b) assuming that $R = 0.025$, the
15   deuterated peptide ($\bullet$) elutes first, and the full peak width at half-maximum ($W_{1/2}$) of the peak is 60 seconds, the deuterated peptide ($\bullet$) would elute 1.5 seconds faster than the nondeuterated peptide ($\blacksquare$), and $ratio_{obs}$ ($\blacklozenge$) would vary continuously across the peak. Eq. 7 approximates the $ratio_{obs}$ well at the center of the peak when R is small ($\blacktriangle$); (c) assuming that $R = 0.5$, the deuterated
20   peptide ($\bullet$) elutes first and the full peak width at half-maximum ($W_{1/2}$) of the peak is 60 seconds, the deuterated peptide ($\bullet$) would elute 30 seconds faster than the nondeuterated peptide ($\blacksquare$), and $ratio_{obs}$ ($\blacklozenge$) would vary continuously across the peak.

Figure 2 depicts a simulation of isotopic fractionation of deuterated ($\bullet$) and
25   nondeuterated ($\blacksquare$) peptides in reversed-phase chromatography, $ratio_{true} = 1$. (a) deuterated and nondeuterated peptides elute at exactly the same time, that is, R = 0. $Ratio_{obs}$ ($\blacklozenge$) does not vary with time and always equals $ratio_{truc}$; (b) R = 0.1, $ratio_{obs}$ ($\blacklozenge$) increases with time and eventually levels off to give $ratio_{true}$.

Figure 3 depicts extracted ion chromatograms of angiotensin I labeled by
30   succinic anhydride and succinic-d_4-anhydride. Four deuterium atoms were incorporated in the deuterated peptide. (a) deuterated peptide ($\bullet$) eluted faster

12

than the nondeuterated peptide (■).   Ratio$_{obs}$ (◆) increased from 0.64 at 25.80

minutes to 1.17 at 26.33 minutes, while ratio$_{true}$ equaled 0.90. Both ratio$_{obs}$ (◆)

vs. time and ln(ratio$_{obs}$) (▲) vs time were linear (R^2 equaled 0.98 and 0.97,

respectively), because the peak asymmetry was small. Both R's calculated from

5   the slopes equaled 0.088 (Eqs. 6 and 8); (b) plotting integrated peak area vs

time, ratio$_{obs}$ (◆) increased with time (i.e., 0.70 at 25.91 minutes) and eventually

leveled off to give ratio$_{true}$ (i.e., 0.90); (c) curve of the dueterated peptide (●)

was normalized (multiplied by ratio$_{true}$) and horizontally shifted toward the right

by 1.94 seconds to make the curves of deuterated (■) and nondeuterated (●)

10   peptides coincide.

Figure 4 depicts extracted ion chromatograms of angiotensin I labeled by

CH_3CO- and CD_3CO- groups. (a) only the NH_2 group at the N-terminus of the

peptide was acetylated. Three deuterium atoms were incorporated into the

deuterated peptide. The deuterated peptide (●) eluted 1.31 seconds faster than

15   the nondeuterated (■) peptide. Ratio$_{obs}$ (◆) increased with time (i.e., 0.76 at

23.80 minutes) and eventually leveled off to give ratio$_{true}$ (i.e., 0.93); (b) both

the NH_2 group at the N-terminus of the peptide and the OH group of tyrosine

were acetylated. Six deuterium atoms were incorporated into the deuterated

peptide. The deuterated peptide (●) eluted 3.75 seconds faster than the

20   nondeuterated peptide (■). Ratio$_{obs}$ (◆) increased with time (i.e., 0.33 at 25.57

minutes) and eventually leveled off to give ratio$_{true}$ (i.e., 0.50).

Figure 5 depicts structures of derivatization agents (a) succinic

anhydride-$^{13}C_0$ and $^{13}C_4$; (b) ICAT-$^2H_0$ and $^2H_8$.

Figure 6 depicts fractionation of a pair of ICAT-$^2H_0$ and $^2H_8$ labeled

25   peptides in reversed-phase chromatography. The peptide sequence is

QNCDQFEK (SEQ ID NO:1). Eight deuterium were incorporated into the

deuterated peptide. Resolution equals 0.74.

(a) Extracted ion chromatogram of the deuterated peptide (▲), non-deuterated

peptide (◆), and ratio between them (■); (b) mass spectrum at 50.7 minutes; (c)

30   mass spectrum at 51.2 minutes; (d) mass spectrum at 51.6 minutes; (e) mass

spectrum of fraction collected at 50.2-50.8 minutes; (f) mass spectrum of

fraction collected at 50.8-51.4 minutes; (g) mass spectrum of fraction collected at 51.4-52.0 minutes.

Figure 7 depicts a pair of succinic anhydride-$^{13}C_0$ and $^{13}C_4$ labeled peptides coelute in revered-phase chromatography. The peptide sequence is

5   QNCDQFEK (SEQ ID NO:1). Eight ^{13}C were incorporated into the heavy isotope labeled peptide. (a) Extracted ion chromatogram of the ^{13}C-labeled peptide ($\blacktriangle$), non-^{13}C-labeled peptide ($\blacklozenge$), and ratio between them ($\blacksquare$); (b) mass spectrum at 28.5 minutes; (c) mass spectrum at 29.0 minutes; (d) mass spectrum at 29.5 minutes.

10   Figure 8 depicts resolution (R) caused by ICAT-$^2H_8$ of tryptic peptides of BSA. Thirteen peptides contain one cysteine and have 8 deuterium incorporated ($\blacklozenge$). Six peptide contain two cysteine and have 16 deuterium incorporated ($\blacksquare$).

Figure 9 depicts resolution (R) caused by succinic anhydride-$^{13}C_4$ on

15   tryptic peptides of BSA. Forty-four peptides contain no lysine and have 4 ^{13}C incorporated ($\blacklozenge$). Twenty-three peptides contain lysine and have 8 13C incorporated ($\blacksquare$).

Figure 10 depicts real-time intelligent data acquisition (IDA) when resolution is negligible. (a) A simulation of abundance ratio ($\blacklozenge$) as a function

20   of elution time in a chromatographic separation of isotopically labeled peptides. R = 0 and ratio ($\blacklozenge$) does not vary with time; (b) A pair of succinic anhydride-$^{13}C_0$ ($\blacklozenge$) and $^{13}C_4$ ($\blacktriangle$) labeled peptides coelute in reversed-phase chromatography. Ratio ($\blacksquare$) is constant in the middle of elution peak where counts are high. Large random errors in ratio ($\blacksquare$) exist at both ends of the

25   elution peak where counts are low; (c) ratios settle quickly when ratios are calculated based on cumulative counts, instead of counts in individual mass spectrum. In this case it reaches accurate ratio at approximately 28.5 minutes, which leaves approximately 1 minute (28.5-29.5 minutes) to perform MS/MS analysis if the ratio exceeds preset threshold.

30   Figure 11 shows mass spectra of a peptide (IFVQK; SEQ ID NO:2) labeled by succinic anhydride (m/z=834.4, control sample), succinic-$^{13}C_2$-anhydride (m/z=838.4, experimental sample 1) and succinic-$^2H_4$-anhydride (m/z

= 842.4, experimental sample 2) at different times: (A) t = 39.025 minutes, (B) t = 39.125 minutes, (C) t = 39.359 minutes.

Figure 12 shows extracted ion chromatograms of the monoisotopic peaks of isotopically labeled peptides in Figure 11 ($\blacktriangle$: m/z=834.4, control

5   sample, $\blacklozenge$: m/z=838.4, experimental sample 1, $\blacksquare$: m/z=842.4, experimental sample 2) and the ratios between experiment and control samples (—: ratio between experimental sample 1 and control sample, $\bullet$: ratio between experimental sample 2 and control sample).

Figure 13 shows the worst cases of each labeling reagent tested in this

10   work. Extracted ion chromatograms of the isotopically labeled peptides ($\blacktriangle$: heavy isotope labeled experimental sample, $\blacklozenge$: non-isotope labeled control sample) and the ratios between experiment and control samples derivatized by various reagents ($\blacksquare$: ratio); (A) Peptide (YIPGTK; SEQ ID NO:3) labeled by propionate-N-hydroxysuccinimide ester and propionate-^{2}H$_5$-N-

15   hydroxysuccinimide ester. Ten ^{2}H were incorporated. R equals 1.0; (B) Peptide (YIPGTK; SEQ ID NO:3) labeled by succinic anhydride and succinic-^{2}H$_4$-anhydride. Eight ^{2}H were incorporated. R equals 0.31; (C) Peptide (IFVQK; SEQ ID NO:2) labeled by N-acetoxysuccinimide and N-acetoxy-^{2}H$_3$-succinimide. Six ^{2}H were incorporated. R equals 0.27; (D) Peptide

20   (TGPNLHGLFGR; SEQ ID NO:4) labeled by H$_2$^{16}O and H$_2$^{18}O. Two ^{18}O were incorporated. R equals 0.061; (E) Peptide (MIFAGIK; SEQ ID NO:5) labeled by succinic anhydride and succinic-^{13}C$_2$-anhydride. Four ^{13}C were incorporated. R equals 0.0096.

Figure 14 shows derivatization of tryptic peptides of cytochrome c by

25   various reagents. Resolution (R) was found to be dependent on peptide mass; (A) Propionate-N-hydroxysuccinimide ester and propionate-^{2}H$_5$-N-hydroxysuccinimide ester ($\blacksquare$: 10 ^{2}H incorporated, $\blacklozenge$: 5 ^{2}H incorporated); (B) Succinic anhydride and succinic-^{2}H$_4$-anhydride ($\blacksquare$: 8 ^{2}H incorporated, $\blacklozenge$: 4 ^{2}H incorporated); (C) N-acetoxysuccinimide and N-acetoxy-^{2}H$_3$-succinimide ($\blacksquare$: 6

30   ^{2}H incorporated, $\blacklozenge$: 3 ^{2}H incorporated); (D) H$_2$^{16}O and H$_2$^{18}O; (E) Succinic

15

anhydride and succinic-$^{13}C_2$-anhydride (■: 4 ^2H incorporated, ◆: 2 ^2H incorporated).

Figure 15 shows the relationship between resolution (R) and various labeling reagents. 1: propionate-$^2H_5$-N-hydroxysuccinimide ester, 2: succinic-$^2H_4$-anhydride, 3: N-acetoxy-$^2H_3$-succinimide, 4: $H_2^{18}O$, 5: succinic-$^{13}C_2$-anhydride. ▲: YIPGTK (SEQ ID NO:3), ■: IFVQK (SEQ ID NO: 2), ◆: TGPNLHGLFGR (SEQ ID NO:4).

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention is directed to materials and methods for minimizing the "isotope effect" during analyte fractionation. Analytes are labeled with non-deuterium isotopes prior to fractionation, resulting in reduction or elimination of the isotope effect.

The invention is especially useful in the area of comparative proteomics, where differential heavy isotope coding methods are used to quantify and identify peptides and proteins in complex mixtures. Minimization of the isotope effect allows the isotope abundance ratio to be measured accurately and quickly, because these ratios can be measured in real time. For example, in comparative proteomics, the invention allows abundance ratios to be determined based on a single mass spectrum taken on-line, in real time, at any point during the reversed phase elution of a complex, isotopically labeled peptide mixture of peptides. Proteins whose abundance ratio differs from the reference ratio can be immediately identified, via MS/MS fragmentation of the peptides. .

The ability to make abundance ratio measurements on-line, in real time, represents a tremendous advance in the art of comparative proteomics. For the first time, chromatographic separation such as reversed phase chromatography (RPC) and mass spectrometric analysis can be executed within a single multidimensional chromatographic system. Upstream reduction, alkylation, proteolysis and/or affinity selection can be added as desired. This advance allows more time to be spent on MS/MS of peptides of interest. The system can even be operated without the attention of an operator, relying on intelligent data acquisition and immediate application of an algorithm that specifies that

16

MS/MS be carried out when the abundance ratio exceeds a predetermined threshold.

Elimination of the isotope effect in accordance with the invention also permits the use of a chemically multiplexed global internal standard technique

5   (mGIST). This technique can be use to concurrently analyze analytes in three, four, five and even more samples, provided that multiple isoforms of the labeling agent are available. mGIST is especially useful in comparative proteomics. For example, it allows discrimination among post-translational variants of proteins, and rapid accommodation of the analysis of structural

10 diversity as described in Example 4.

The identical or nearly identical chromatographic retention times of analyte isoforms also serves as one of the criteria to recognize isoform pairs from among multiple analytes with equivalent masses observed using MS. Proteomes are extremely complicated, and many totally irrelevant peptides

15 could appear to be isoform pairs based on mass alone. Without a means to distinguish isoform pairs from other species having the same mass, abundance ratio is calculated and an up/down regulation may be erroneously reported. This kind of erroneous conclusions can be avoided if additional data is available that might otherwise distinguish the species, such as chromatographic retention

20 time, MS/MS spectra, etc.

Isotope effect

For mass spectrometric analysis of comparative samples as described above, it is desired that the labeled analytes be chemically equivalent but

25 isotopically distinct. Labeled analytes that are chemically equivalent but isotopically distinct are referred to as isotopic isoforms or simply isoforms. Isotopic isoforms are "chemically equivalent" in that they have the same chemical composition and structure; however they are "isotopically distinct" in that at least one atom in a first isoform is substituted with a heavy isotope of

30 that atom in a second isoform. It is evident that many different isoforms of a compound can be synthesized, depending on the number and type of atoms substituted with a heavy isotope of that atom.

17

Because they are chemically equivalent, isotopic isoforms typically exhibit identical or substantially identical behavior during physical or chemical fractionation processes, such as chromatography or electrophoresis, such that isoforms cannot be readily separated from one another using standard laboratory

5   purification and separation techniques. For example, a protein or peptide present in each sample may, after labeling, differ in mass by a few atomic mass units when the protein or peptide from one sample is compared to the same protein or peptide from the other sample (i.e., they are isotopically distinct). However, these two proteins or peptides would ideally exhibit identical

10   chromatographic behavior and electrophoretic migration patterns.

To the extent isotopic isoforms exhibit non-identical fractionation behavior, this is referred to as the "isotope effect." Unless otherwise indicated, the term "isotope effect," when used herein to describe the behavior of a heavy atom isoform of a compound, means the difference in fractionation behavior of

15   the heavy atom isoform and an isoform of that compound containing no heavy atoms.

A primary objective of the present invention is to reduce or eliminate the isotope effect. Isotope coding agents that contain heavy isotopes other than deuterium in accordance with the invention can be employed to reduce or

20   eliminate the "isotope effect" that characterizes deuterated isoforms.

Labeling agents and non-deuterium mass isotopes

It has been discovered that the isotope effect is ameliorated when heavy isotopes other than deuterium are used to isotopically code peptides and

25   proteins. As noted in more detail below, this strategy can be used with amino acid specific coding agents, such as the -SH labeling agent known by the tradename "ICAT" (isotope coding affinity tag) (Gygi et al., Nat. Biotechnol. (1999) 17, 994-999), or with the global internal standard technique (GIST) (PCT WO 01/86306, published November 15, 2001, Ji et al., *J. Chromatogr. B,*

30   (2000) 745, 197-210). Examples of stable mass isotopes (other than deuterium) that can be used to label an analyte include ^{13}C, ^{15}N, ^{17}O, ^{18}O, ^{33}S, and ^{34}S but should be understood that the invention is in no way limited by the choice of

18

isotope. For purposes of the invention, the term "non-deuterium heavy isotope" includes all stable mass isotopes other than ^{2}H.

A heavy isotope can be incorporated into a labeling agent such as an affinity tag, or it can be linked to the peptide or protein in a separate chemical or enzymatic reaction. The term "isotope coding agent" refers to a reagent in the form of, for example, a labeling agent, such as ICAT, containing one or more heavy isotopes, which can be covalently linked to an analyte to isotopically label it. A "labeling agent" is a reagent includes a reactive group capable of covalent linkage to the analyte, such as a peptide or protein, and, optionally, an affinity tag to facilitate selection of the labeled analytes. Affinity selection of peptides is an optional step the comparative analysis of biological metabolites (see below), thus the inclusion of an affinity ligand in the labeling agent is optional.

The term "isotope coding agent" also includes a heavy isotope that is incorporated into a metabolite either post-synthetically, such as ^{18}O that is incorporated into proteins and peptides by isotope exchange with $H_2{}^{18}$O during proteolysis or deglycosylation, or during biosynthesis, such as by providing a cell with enriched media containing ^{13}C- or ^{15}N- labeled amino acids during protein synthesis.

Any labeling agent with known or expected utility as a deuterated isotope coding agent can be employed as an isotope coding agent containing ^{13}C, ^{18}O and/or ^{15}N instead of deuterium, provided the labeling agent contains carbon, oxygen and/or nitrogen atoms, respectively. Examples of labeling agents include succinic anhydride and ICAT reagent or derivative thereof.

Biological samples

Samples analyzed according to the method of the invention are preferably obtained from a "biological environment," which is to be broadly interpreted to include any type of biological system in which enzymatic reactions can occur, including *in vitro* environments, cell culture, cells at any developmental stage, whole organisms, organs, tissues, bodily fluids, and the like.

19

Metabolites such as proteins (or peptides if proteolysis is employed) in control and experimental samples are labeled (either post-synthetically or during biosynthesis) with distinct isotopic forms of a labeling agent. In proteomics, MS can be performed on polypeptides of any length. Proteins are
5 typically fragmented into peptides prior to labeling, but whole proteins can be labeled if desired.

The terms "polypeptide," "protein/peptide" and "protein" as used herein are equivalent and include both peptides (i.e., short polypeptides, typically less than about 50 amino acids, more typically less than about 30 amino acids), and
10 longer polypeptides. These terms are used interchangeably. Furthermore, unless otherwise indicated, techniques described with reference to longer polypeptides are equally applicable to shorter polypeptides (peptides), and vice versa. It should be understood that the terms "polypeptide," "protein/peptide" and "protein" refer to a polymer of amino acids and do not connote a specific
15 length of a polymer of amino acids. Thus, for example, the terms oligopeptide, polypeptide, and enzyme are included within the definition of polypeptide, protein/peptide and protein, whether produced using recombinant techniques, chemical or enzymatic synthesis, or naturally occurring. These terms also includes polypeptides that have been modified or derivatized, such as by
20 glycosylation, acetylation, phosphorylation, and the like. When the term "peptide" is used herein, it generally refers to a protein fragment produced in solution.

The method of the invention is especially well-suited for use in complex samples containing a number of different metabolites, although it is applicable
25 to less complex samples as well. In comparative proteomics applications, the sample preferably contains at least about two proteins; more preferably it contains at least about 100 proteins; still more preferably it contains at least about 1000 proteins. A sample can therefore include total cellular protein or some fraction thereof. For example, a sample can be obtained from a particular
30 cellular compartment or organelle, using methods such as centrifugal fractionation. The sample can be derived from any type of cell, organism, tissue, organ, or bodily fluid, without limitation. The method of the invention can be used to identify one or more proteins in the sample, and is typically used

20

to identify multiple proteins in a single complex mixture. It should therefore be understood that when the method of the invention is referred to, for simplicity, as a method for identifying "a protein" in a mixture that contains multiple proteins, the term "a protein" is intended to mean "at least one protein" and thus

5 includes one or more proteins.

Comparative analysis of biological metabolites

There is a growing need to move beyond the massive effort to define genetic and protein components of biological systems to the study of how they

10 and other cellular metabolites are regulated and respond to stimuli. Advantageously, the invention permits faster and more accurate identification, compared to previous methods, of up- and down-regulated metabolites in biological samples in response to the application of a stimulus. The words "stimulus" and "stimuli" are used broadly herein and mean any agent, event,

15 change in conditions or even the simple passage of time that may be associated with a detectable change in expression of at least one metabolite within a cell, without limitation. For example, a stimulus can include a change in growth conditions, disease state, pH, nutrient supply, temperature, electrical or electrochemical environment, or pressure; contact with an exogenous agent such

20 as a drug or microbe, competition with another organism; and the like. The term "metabolite" refers, in this context, to a cellular component, preferably an organic cellular component, which can change in concentration in response to a stimulus, and includes large biomolecules such as proteins, polynucleotides, carbohydrates and fats, as well as small organic molecules such as hormones,

25 peptides, cofactors and the like.

Advantageously, in comparative analysis of biological metabolites according to the invention, it is not necessary to preselect the change in concentration that will be studied; instead, any of the hundreds of thousands or millions of the proteins in the samples can change in relative concentration, and

30 these proteins can be readily identified.

As noted above, labeling of biological metabolites using isotope coding agents that contain heavy isotopes other than, or in addition to, deuterium in

accordance with the invention reduces or eliminates the "isotope effect" observed during fractionation of deuterated samples.

Isotopically labeled samples are combined, and the mixture is then fractionated using one or more selection or fractionation techniques. The

5 objective of fractionation is to reduce sample complexity to the extent that abundance ratio analysis can be performed, using a mass spectrometer, on individual analyte pairs. The method is not limited by the techniques used for selection and/or fractionation. Typically, fractionation is carried out using single or multidimensional chromatography such as reversed phase

10 chromatography (RPC), ion exchange chromatography, hydrophobic interaction chromatography, size exclusion chromatography, or affinity fractionation such as immunoaffinity and immobilized metal affinity chromatography. Preferably the fractionation involves surface-mediated selection strategies. Electrophoresis, either slab gel or capillary electrophoresis, can also be used to

15 fractionate the analytes. Examples of slab gel electrophoretic methods include sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and native gel electrophoresis. Capillary electrophoresis methods that can be used for fractionation include capillary gel electrophoresis (CGE), capillary zone electrophoresis (CZE) and capillary electrochromatography (CEC), capillary

20 isoelectric focusing, immobilized metal affinity chromatography and affinity electrophoresis.

Because >95% of cellular proteins do not change in response to a stimulus, proteins (as well as other metabolites) in flux can be readily identified by abundance ratio changes in species resolved, for example, by 2-D gel

25 electrophoresis or liquid chromatography. Once these proteins are detected, they can optionally be identified using the "signature peptide" approach as described in (PCT WO 01/86306, published November 15, 2001, Ji et al., *J. Chromatogr. B*, (2000) 745, 197-210) or any other convenient method. One example of how this method of the invention can be used is to analyze patterns

30 of protein expression in a breast cancer cell before and after exposure to a candidate drug. The method can also be used to analyze changes in protein expression patterns in a cell or an organism as a result of exposure to a harmful agent. As yet another example, the method can be used to track the changes in

22

protein expression levels in a cell as it is exposed, over time, to changes in light, temperature, electromagnetic field, sound, humidity, and the like.

Post-synthetic isotope labeling of proteins advantageously creates internal standards from proteins of unknown structure and concentration.

5   Whenever there is a control, or reference state, in which the concentration of proteins is at some reference level, proteins in this control state can serve as internal standards. In a preferred embodiment of the invention, constituent peptides are labeled after proteolytic cleavage proteins in the sample. The timing of the labeling step provides an opportunity to label every peptide in the

10   mixture by choosing a labeling method that labels at the N or the C terminus of a polypeptide, in what is referred to as the "Global Internal Standard Technique" or "GIST". The GIST strategies exploit the facts that 1) proteolysis generates primary amine and carboxyl groups during peptide bond hydrolysis, 2) primary amines are easily acetylated with N-acetoxysuccinimide, and 3) ^{18}O

15   is easily incorporated into carboxyl groups during proteolysis. However, it should be noted that the resolution of some ^{18}O labeled peptides was larger than with ^{2}H labeled coding reagents.

Isotopic labeling after the proteins have been synthesized has a further advantage. Although metabolic incorporation of labeled amino acids has been

20   widely used to label proteins and is not to be excluded from the invention, it is not very reproducible and is objectionable in human subjects. Post-sampling strategies for incorporation of labels are much more attractive.

An advantage of the internal standard method is that it detects relative changes in, not changes in absolute amounts of, analytes. It is very difficult to

25   determine changes in absolute amounts analytes that are present at very low levels. This method is as sensitive to changes in very dilute analytes as it is those that are present at great abundance. Another important advantage of this approach is that it is not influenced by quenching in the MALDI. This means that large number of peptides can be analyzed irrespective of the expected

30   quenching.

In comparative proteomics applications, it is often advantageous to cleave proteins in a sample into their constituent peptides at some point prior to mass spectrometric analysis. Fragmentation of proteins can be achieved by

23

chemical, enzymatic or physical means, including, for example, sonication or shearing. Preferably, a protease enzyme is used, such as trypsin, chymotrypsin, papain, gluc-C, endo lys-C, proteinase K, carboxypeptidase, calpain, subtilisin and pepsin; more preferably, a trypsin digest is performed. Alternatively,

5   chemical agents such as cyanogen bromide can be used to effect proteolysis. The proteolytic agent can be immobilized in or on a support, or can be free in solution.

In one embodiment of the method of the invention, proteins are isotopically labeled prior to cleavage. For example, proteins in a control sample can be derivatized with a labeling agent that contains a heavy isotope (an

10   isotope coding agent), while proteins in an experimental sample can be derivatized with the normal labeling agent. The samples are then combined for further analysis. The derivatized proteins can be chemically or enzymatically cleaved, if desired, either before or after fractionation. Cleavage is optional;

15   isotopically labeled proteins can, if desired, be analyzed directly following a fractionation step such as multidimensional chromatography, 2-D electrophoresis or affinity fractionation.

When the derivatized proteins are cleaved before fractionation, the labeling agent preferably contains an affinity ligand, and the tagged peptide

20   fragments are first affinity selected, then fractionated in a 1-D or 2-D chromatography system, after which they are analyzed using mass spectrometry (MS). In instances where the derivatized proteins are cleaved after fractionation, 2-D gel electrophoresis is preferably used to separate the proteins. If the peptides have also been affinity labeled, selection of the affinity-tagged peptides

25   can be performed either before or after electrophoresis.

Mass spectrometric analysis can be used to determine peak intensities and quantitate abundance ratios in the combined sample, determine whether there has been a change in the concentration of a protein between two samples, and to facilitate identification of a protein from which a peptide fragment,

30   preferably a signature peptide, is derived. Preferably, changes in peptide concentration between the control and experimental samples are determined by abundance ratio MALDI-mass spectrometry because MALDI-MS allows the

24

analysis of more complex peptide mixtures, but ESI-MS may also be used when the peptide mixture is not as complex.

In a complex combined mixture, there may be hundreds to thousands of peptides, and many of them will not change in concentration between the

5   control and experimental samples. These peptides whose levels are unchanged are used to establish the normalized abundance ratio for peptides that were neither up nor down regulated. All peptides in which the abundance ratio exceeds this value are up regulated. In contrast, those in which the ratio decreases are down regulated. A difference in relative abundance ratio of a

10   peptide pair, compared to peptide pairs derived from proteins that did not change in concentration, thus signals a protein whose expression level did change between the control and experimental samples. If the peptide characterized by an abundance ratio different from the normalized ratio is a signature peptide, this peptide can be used according to the method of the

15   invention to identify the protein from which it was derived.

If desired, isotope labeling can take place after cleavage of the proteins in the two samples. Derivatization of the peptide fragments is accomplished using a labeling agent that preferably contains an affinity ligand. On the other hand, an affinity ligand can be attached to the peptides in a separate reaction,

20   either before or after isotopic labeling. If attached after isotopic labeling, the affinity ligand can be attached before or after the samples are combined. The peptide fragments in the combined mixture are affinity selected, then optionally fractionated using a 1-D or multi-dimensional chromatography system, or a capillary or slab gel electrophoretic technique, after which they are analyzed

25   using mass spectrometry. In instances where the peptides are not affinity tagged, they are either affinity selected based on their inherent affinity for an immobilized ligand (preferably using IMAC or immobilized antibody or lectin) or analyzed without selection.

30

Post-synthetic isotope labeling techniques

1. *Alkylation with isotopically distinct reagents*

It is often desirable to protect free thiols by alkylation after reduction
5 and prior to fractionation and mass spectrometry. Accordingly, proteins or
other thiol-containing metabolites in control and experimental samples are
optionally alkylated using alkylating agents. The alkylating agent can double as
an isotope coding agent, in which case different isoforms of the alkylating agent
are used for control and experimental samples.

10   Iodoacetic acid (ICH_2COOH) is an example of an alkylating agent.
Normal iodoacetate can be used to derivatize the control, for example, and ^{13}C
iodoacetate can be used to derivatize the experimental sample. Preferably, the
^{13}C isoform contains two atoms of ^{13}C.

Based on the fact that proteins from control and experimental samples
15 are identical in all respects except the isotopic content of the iodoacetate
alkylating agent, their relative molar response (U) is expected to be 1. This has
several important ramifications. When control and experimental samples are
mixed:

$$A = \Lambda \Delta$$

20

In this case Δ will be i) the same for all the proteins in the mixture that do not
change concentration in the experimental sample and ii) a function of the
relative sample volumes mixed. If the protein concentration in the two samples
is the same and they are mixed in a 1/1 ratio for example, then $\Delta = 1$. With a
25 cellular extract of 20,000 proteins, Δ will probably be the same for >19,900 of
the proteins in the mixture. The concentration of a regulated protein that is
either up- or down-regulated is expressed by the equation:

$$A_{expt.} = \Lambda_{cont.} \Delta \delta$$

30

where $A_{expt.}$ is a protein from the experimental sample that has been
synthetically labeled with a derivatizing agent, $\Lambda_{cont.}$ is the same protein from
the control sample labeled with a different isotopic form of the derivatizing

agent, and δ is the relative degree of up- or down-regulation. Because Δ is an easily determined constant derived from the concentration ratio of probably >95% of the proteins in a sample, δ is readily calculated and proteins in regulatory flux easily identified.

5   Another example of an alkylating agent that is useful for isotope coding is an ICAT (isotope coding affinity tag) reagent. An ICAT reagent includes a reactive functional group that reacts with an amine, a thiol, a hydroxyl or a carboxyl; an isotopic linker; and, optionally, an affinity functional group for affinity selection. The "linker" links the reactive functional group and the

10   optional affinity functional group. In a preferred embodiment, the reactive function group is reactive group such as iodoacetic acid or iodocetamide, capable of covalent linkage to a thiol-containing metabolite such as a protein or peptide.

          For example, an ICAT reagent can alkylate the free cysteines in a

15   protein or peptide. An affinity tag, such as biotin, facilitates selection of the tagged compounds. To function as an isotope coding agent, an ICAT reagent further includes an isotopically labeled linker, e.g., $C_{10}H_{17}N_3O_3$, which can carry up to ten ^{13}C atoms, three ^{18}O atoms or three ^{15}N atoms, or any combination thereof. An ICAT reagent containing a linker having nine ^{13}C

20   isotopes is a particularly preferred embodiment of the invention and can be obtained from Applied Biosystems, Inc.

          The ICAT reagent of the invention is very versatile isotope coding agent and is expected to be useful in experiments utilizing the multiplexed global internal standard technique (mGIST) as described in more detail below.

25   It should be understood that the ICAT reagent is not limited by the particular isotopic linker, and that linkers other than $C_{10}H_{17}N_3O_3$ can be used. Optionally, the ICAT reagent also includes a cleavage site that allows removal of the affinity portion of the label and part of the linker reducing the overall mass of the tag on the peptides and improving the overall peptide fragmentation

30   efficiency.

2. *Isotopic labeling of amines*

If not included as part of the alkylating agent, an isotope label can be applied to the peptide or other metabolite as part of an affinity tag (if affinity selection is contemplated), or at some other reactive site on the peptide.

5   Although application of the internal standard isotopic label in the affinity tag is operationally simpler and, in some cases, more desirable, it requires that each affinity tag be synthesized in at least two isotopic forms. Amine-labeling in a separate step (i.e., uncoupling the label and the affinity ligand) is therefore a preferred alternative.

10   Peptides that are generated by trypsin digestion (as well as those generated by many other types of cleavage reactions) have a primary amino group at their amino-terminus in all cases except those in which the peptide originated from a blocked amino-terminus of a protein. Moreover, the specificity of trypsin cleavage dictates that the C-terminus of signature peptides

15   will have either a lysine or arginine (except the C-terminal peptide from the protein). In rare cases there may also be a lysine or arginine adjacent to the C-terminus.

Primary amino groups are easily acylated with, for example, acetyl N-hydroxysuccinimide (ANHS). Indeed, any analyte having a carboxyl group

20   can be activated by NHS then derivatized with a labeling agent that having a free amine. Control samples can be acetylated with normal ANHS whereas experimental tryptic digests can be acylated with either $^{13}CH_3CO\text{-}NHS$ or other ^{13}C-containing derivatizing agent. Our studies show that the ε-amino group of all lysines can be derivatized in addition to the amino-terminus of the peptide,

25   as expected. This is actually an advantage in that it allows a determination of the number of lysine residues in the peptide.

28

Essentially all peptides in both samples will be derivatized and hence distinguishable from their counterparts using mass spectrometry. This means that any affinity selection method or combination of affinity selection methods (other than possibly those that select for arginine or lysine, which contain free amines) can be used at any point in the process to obtain a selected population enriched for signature peptides. For example, isotope labeling at amines can be used to identify changes in the relative amounts of peptides selected on the basis of cysteine, tryptophan, histidine, and a wide variety of post-translational modifications. In this preferred embodiment of the method, isotopic labeling and affinity labeling are two independent and distinct steps, and virtually all peptides are isotopically labeled. This provides significantly more flexibility and greater control over the production of signature peptides than is possible when the alkylating agent doubles as the isotope labeling agent.

3. *Isotopic labeling of hydroxyls and other functional groups*

While acetylation is a convenient labeling method for proteins and their constituent peptides, other labeling methods may be useful for other types of cellular metabolites. For example, acetic anhydride can be used to acetylate hydroxyl groups in the samples, and trimethylchlorosilane can be used for less specific labeling of functional groups including hydroxyl groups, carboxylate groups and amines.

Mass spectrometry

Masses of the fractionated metabolites are preferably determined by mass spectrometry, preferably using matrix assisted laser desorption ionization (MALDI) or electrospray ionization (ESI), and mass of the peptides is analyzed using time-of-flight (TOF), quadrapole, ion trap, magnetic sector or ion cyclotron resonance mass analyzers, or a combination thereof including, without limitation, TOF-TOF and other combinations. The mass of peptides analyzed in comparative proteomics applications is preferably determined with a mass accuracy of about 10 ppm or better; more preferably, masses are determined with a mass accuracy of about 5 ppm or better; most preferably they are determined with a mass accuracy of about 1 ppm or better. The lower the

29

ppm value, the more accurate the mass determination and the less sequence data
is needed for peptide identification.

Interpretation of the spectra

5   Isotopic isoforms of a metabolite from control and experimental samples
that have been mixed together will exhibit identical, or nearly identical,
behavior during fractionation prior to mass spectrometry and will not be
resolved. A mass spectrometer, however, can readily differentiate between
species labeled with a heavy isotope and normal species. In proteomics
10   applications, mass spectometric analysis of isoforms present as either as
proteolytic fragments or as a low molecular weight (e.g., under about 15kD)
proteins allows ratios of protein abundance ("abundance ratios") between the
two samples to be established. The relative abundance of most proteins will be
the same and allow Δ to be calculated. A second group of proteins will be seen
15   in which the relative abundance of specific proteins is much larger in the
experimental sample. These are the up-regulated proteins. In contrast, a third
group of proteins will be found in which the relative abundance of specific
proteins is lower in the experimental sample. These are the down-regulated
proteins. The degree (δ) to which proteins are up- or down-regulated is
20   calculated based on the computed value of Δ.

A more detailed analysis of the interpretation of the resulting mass
spectra is provided using amine-labeled proteins as an example. Signature
peptides of experimental samples in this example are acetylated at the amino-
termini and on ε-amino groups of lysines with either $^{13}CH_3CO$- or CD_3CO-
25   residues, therefore any particular peptide will appear in the mass spectrum as a
doublet. Although this example employs deuterium as the heavy isotope, it is
readily applicable to isotope labeling with non-deuterium heavy isotopes in
accordance with the invention, to reduce the isotope effect in fractionation
procedures prior to mass spectrometry.

30   In the simplest case, which involves the use of deuterium as the heavy
and is thus presented here for illustrative purposes only, where i) trideutero-
acetic acid is used as the labeling agent, ii) the C-terminus is arginine, iii) there
are no other basic amino acids in the peptide, and iv) the control and

30

experimental samples are mixed in exactly a 1/1 ratio before analysis, i.e., $\Delta=1$, the spectrum shows a doublet with peaks of approximately equal height separated by 3 amu. With 1 lysine the doublet peaks were separated by 6 amu and with 2 lysine by 9 amu. For each lysine that is added the difference in mass

5   between the experimental and control would increase an additional 3 amu. It is unlikely in practice that mixing would be achieved in exactly a 1/1 ratio. Thus Δ will have to be determined for each sample and varies some between samples. Within a given sample, Δ will be the same for most peptides, as will also be the case in electrophoresis. Peptides that deviate to any extent from the average

10   value of Δ are the ones of interest. The extent of this deviation is the value δ, the degree of up- or down-regulation. As indicated above, Δ will be the same for greater than 95% of the proteins, or signature peptides in a sample.

As noted above, amino acids with other functional groups are occasionally labeled. In the presence of a large excess of acylating agent

15   hydroxyl groups of serine, threonine, tyrosine, and carbohydrate residues in glycoconjugates and the imidazole group of histidine can also be derivatized. This does not interfere with quantification experiments, but complicates interpretation of mass spectra if groups other than primary amines are derivatized. In the case of hydroxyl groups, esters formed in the derivatization

20   reaction are readily hydrolyzed by hydroxylamine under basic conditions. Aclylation of imadazole groups on the other hand occurs less frequently than esterification and is perhaps related to amino acid sequence around the histidine residue.

Another potential problem with the interpretation of mass spectra in the

25   internal standard method of the invention can occur in cases where a protein is grossly up- or down-regulated. Under those circumstances, there will essentially be only one peak. When there is a large down-regulation this peak will be the internal standard from the control. In the case of gross up-regulation, this single peak will have come from the experimental sample. The

30   problem is how to know whether a single peak is from up- or down-regulation. This is addressed by double labeling the control with $CH_3CO\text{-}NHS$ and $^{13}CH_3CO\text{-}NHS$. Because of the lysine issue noted above, it is necessary to split the control sample into two lots and label them separately with $CH_3CO\text{-}NHS$

31

and $^{13}CH_3CO$-NHS, respectively, and then remix. When this is done the control always appears as a doublet separated by 1-2 amu, or 3 amu in the extreme case where there are two lysines in the peptide. When double labeling the control with ^{12}C and ^{13}C acetate and the experimental sample with trideuteroacetate,

5   spectra would be interpreted as follows. A single peak in this case would be an indicator of strong up-regulation. The presence of the internal standard doublet alone would indicate strong down-regulation.

Another potential problem with the double labeled internal standard is how to interpret a doublet separated by 3 amu. Because the control sample was

10   labeled with CH_3CO-NHS and $^{13}CH_3CO$-NHS, this problem can arise only when the signature peptide has 2 lysine residues and is substantially down-regulated to the point that there is little of the peptide in the experimental sample. The other feature of the doublet would be that the ratio of peak heights would be identical to the ratio in which the isotopically labeled control peptides

15   were mixed. Thus, it may be concluded that any time a doublet appears alone in the spectrum of a sample and Δ is roughly equivalent to that of the internal standard that i) the two peaks came from the control sample and ii) peaks from the experimental sample are absent because of substantial down regulation.

20   *Microfabricated analytical systems*

The method of the invention is amenable to automation by integrating most of the analytical steps in a single instrument. Reduction, alkylation, proteolysis, affinity selection, and reversed phase chromatography (RPC), or any subset thereof, can be executed within a single multidimensional

25   chromatographic system. Samples collected from this system are manually transferred to MALDI plates for mass spectrometric analysis. In one embodiment, the invention provides a single channel integrated system. In a preferred embodiment, however, the invention thus provides a microfabricated, integrated, parallel processing, microfluidic system that carry out all the

30   separation components of analysis on a single chip.

Real-time intelligent data acquisition and analysis (IDA)

Proteolysis of a proteome produces very large numbers of peptides and identifying all these peptides by MS/MS is a lengthy, formidable task. However, studies have shown that only a small percentage of all the proteins in

5   a proteome change significantly in concentration as a result of some stimulus. When the objective it to identify only those proteins that have undergone change (comparative proteomics), throughput and data quality could be greatly improved by fragmenting only those peptides of interest. Real-time abundance ratio analysis would allow a mass spectrometer (MS) and data system to select

10   peptides for MS/MS sequencing based on changes in relative concentration that exceed preset values.

Minimizing the isotope effect by controlling "R", the difference in elution times between isoforms, using the techniques described herein allows "real time ratio analysis" so that decisions regarding further (e.g., MS) analysis

15   of chromatographic peaks can be made in real time.   R (resolution) is defined as the differences between the retention times divided by the average peak width at half maximum.

$$R = \frac{t_2 - t_1}{w_{1/2}}$$

20   Further analysis is initiated only on the peptides or proteins whose changes (as evidenced by the abundance ratio) exceed the preset values. With real-time abundance ratio analysis, the amount of data that needs to be analyzed is substantially reduced, and throughput can be doubled.

There are a number of ways to do "real time ratio analysis". Using ^{13}C

25   coding agents, which essentially exhibit no isotope effect, the abundance ratio does not change significantly across the LC elution peak, so the ratio can be determined by a single snapshot in MS. Although isotopically labeled isoforms of peptides are not resolved in the case of ^{13}C coding, large random errors in abundance ratio are often seen at both the leading and tailing edges of eluting

30   peaks where concentration and counts are low.   In the middle of a peak, however, counts are high. It is often desirable, however, to determine the abundance ratio as early as possible, preferably prior to the elution of the middle

33

of the peak. One way to improve the accuracy of an early abundance ratio determination is to carry out the following steps.

First, identify pairs of isotopically labeled peptides (isoform pairs) based on the mass differences, similar peak profile, constant abundance ratio across

5   time, etc. Using non-deuterium isotopes such as 13C helps to identify of isoform pairs correctly, because the retention time of the isoforms should be the same. This means that the abundance ratio should be constant across the elution peak. If it is observed that the retention times are not the same, or on the other hand the abundance ratio is very different across the peak, it means the

10   suspected isoform pair may not be a true pair, or perhaps that one of them is overlapped with other peptides, and the pair is not "pure." Second, trace accumulative counts of both isotopically labeled peptides. Third, average the ratio between the accumulated counts of subsequent measurements when they reach a preset level. Finally, perform an MS/MS analysis when the ratio

15   exceeds the preset threshold. An example is seen in Figure 10C of Example II. The ratio could be determined accurately at 28.5 minutes, which leaves approximately 1 minute for further analysis of the peptides if the ratio exceeds a preset threshold. The ratio could be determined even earlier if the intensities of the peptides are higher than in Figure 6.

20   For reagents that produce a small isotope effect (e.g. ^{18}O), the elution peaks of the isotopically labeled peptides can be tracked individually, and once they both reach the peak maximum, the first half of the elution peaks (from the start of the peak to the peak maximum) are integrated, the abundance ratio is calculated based on the integrated areas, and a decision is made whether further

25   analysis need to be performed by MS.

Although abundance ratio would be the primary selection criterion, difference in mass between ion clusters could also be used. For example, peptides labeled with succinate in Example II varying by 4 amu contained a C-terminal arginine while those varying by 8 amu contained a C-terminal lysine.

30   As more coding agents become available, it is likely that different types of derivatization will be coded by this mass shift technique.

Real-time IDA increases throughput, reduces sample consumption, and improves the quality of MS/MS data because more time can be spent on the

34

peptides of interest. Isotopically labeled peptides can be fragmented together when resolution is small, which could help to identify the y and b ion series. For example, y ions are paired with mass differences of 4 Da and b ions are unpaired when the carboxyl terminus of the tryptic peptides are labeled by ^{18}O

5   during trypsin digestion.

Software development

Software currently in use in some commercially available instruments instructs mass spectrometers to perform real-time MS-MS on the species that

10   give the most intense signals in mass spectrum. Typically the top 3 signals are analyzed. The software directs selective fragmentation of the most intense isoform pairs, and ignores all the singlets. This scheme reduces the number of unnecessary MS/MS experiments by 50%. However, the most intense signal doesn't necessarily mean that species has the highest concentration. Moreover,

15   those metabolites that are most abundant may not be the ones that have significant change in relative concentration.

The present invention allows real-time MS/MS to be performed on those isoforms that evidence a change in relative concentration, regardless of abundance as long as the abundance is sufficient to rise above the level of

20   random error. By focusing on changed ratios, which include the singlets, it is expected that MS/MS experiments can be reduced by about 80%.

The isotope labeling method of the invention allows the identification of the small number of proteins (peptides) in a sample that are in regulatory flux. Observations of spectra with 50 or fewer peptides indicate that individual

25   species generally appear in the spectra as bundles of peaks consisting of the major peptide ion followed by the ^{13}C isotope peaks. Once a peak bundle has been located, peak ratios within that bundle are evaluated and compared with adjacent bundles in the spectrum. Based on the isotopes used in labeling, simple rules can be articulated for the identification of up- and down-regulated

30   peptides in mass spectra. Software can be written that apply these rules for interpretation.

Data processed in this way can be evaluated in several modes. One is to select a given peptide and then locate all other peptides that are close in δ value.

All peptides from the same protein should theoretically have the same δ value (i.e., the same relative degree of up- or down-regulation). For example, when more than one protein is present in the same 2-D gel spot there is the problem of knowing which peptides came from the same protein. The δ values are very

5   useful in this respect, and provide an additional level of selection. The same is true in 2-D chromatography. 3-D regulation maps of chromatographic retention time vs. peptide mass vs. δ can also be constructed. This identifies proteins that are strongly up- or down-regulated without regard to the total amount of protein synthesized. In some experiments, one or more groups of proteins may be

10   identified that have similar δ values, and identification of the members of a group may elucidate metabolic pathways that had not previously been characterized.

Construction of temporal maps

15   The discussion above would imply that regulation is a process that can be understood with single measurements, i.e., after a stimulus has been applied to a biological system one makes a measurement to identify what has been regulated. However, single measurements at the end of the process only identify the cast of characters. Regulation involves adjusting, directing,

20   coordinating, and managing these characters. The issue in regulation is to understand how all these things occur. Regulation is a temporal process involving a cascade of events. Consider, for example, the hypothetical case in which an external stimulus might cause modification of a transcription factor, which then interacts with another transcription factor, the two of which initiate

25   transcription of one or more genes, which causes translation, and finally post-translational modification to synthesize another transcription factor, etc. Temporal analysis brings a lot to understanding this process. Global analysis of protein synthesis in response to a variety of stimuli has been intensely examined and at least two mapping strategies have been developed (R. VanBogelen et al.,

30   in F. Neidhardt et al., Ed. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd Ed. ASM Press, Washington D.C. , pp. 2067-2117); H. Zhang et al., J. Mass Spec. 31:1039-1046 (1996)).

A temporal map of protein expression can be constructed by first identifying all species that change in response to a stimulus, then performing a detailed analysis of the regulatory process during protein flux. Identification of those proteins affected by the stimulus is most easily achieved by a single

5   measurement after the regulatory event is complete and everything that has changed is in a new state of regulation. Both chromatographic and electrophoretic methods can be used to contribute to this level of understanding. The regulatory process during protein flux is then analyzed at short time intervals and involves many samples. The initial identification process yields

10   information on which species are in flux, their signature peptides, and the chromatographic behavior of these peptides. As a result, the researcher thus knows which samples contain specific signature peptides and where to find them in mass spectra. Quantitating the degree to which their concentration has changed with the internal standard method is straightforward. The resulting

15   data allows temporal maps of regulation to be constructed, and the temporal pattern of regulation will provide information about the pathway of response to the stimulus. The invention thus further provides a method for developing algorithms that identify signature peptides in regulatory change.

20   *Affinity selection*

Affinity selection, although optional, is useful to help identify which proteins in a complex mixture have undergone up- or down-regulation as a result of the application of a stimulus. Preferably, an affinity selection process is used to select peptide fragments that contain specific amino acids, thereby

25   substantially reducing the number of sample components that must be subjected to further analysis. The affinity selection can be effected using an affinity ligand that has been covalently attached to the protein (prior to cleavage) or its constituent peptides (after cleavage), or using an endogenous affinity ligand.

The affinity selection is preferably based on low abundance amino acids

30   or post-translational modifications so as to preferentially isolate "signature peptides." Peptides from complex proteolytic digests that contain low abundance amino acids or specific post-translational modifications are selected (purified) to reduce sample complexity while at the same time aiding in the

37

identification of peptides selected from the mixture. Selection of peptide fragments that contain cysteine, tryptophan, histidine, methionine, tyrosine, tyrosine phosphate, serine and threonine phosphate, O-linked oligosaccharides, or N-linked oligosaccharides, or any combination thereof can be achieved. It is

5   also possible to determine whether the peptide has a C-terminal lysine or arginine and at least one other amino acid.

The method is not limited by the affinity selection method(s) employed and nonlimiting examples of affinity selections are described herein and can also be found in the scientific literature, for example in M. Wilchek, Meth.

10   Enzymol. 34, 182-195 (1974). This approach enormously reduces the complexity of the mixture. If desired, two or more affinity ligands (e.g., primary and secondary affinity ligands) can be used, thereby allowing a finer selection.

Preferably, the affinity selected peptides are subjected to a fractionation

15   step to further reduce sample size prior to the determination of peptide masses. A premise of the signature peptide strategy is that many more peptides are generated during proteolysis than are needed for protein identification. This assumption means that large numbers of peptides potentially can be eliminated, while still leaving enough for protein identification.

20

Affinity tags

An affinity tag used for selection can be endogenous to the protein, or it can be added by chemical or enzymatic processes. The term "affinity tag," as used herein, refers to a chemical moiety that functions as, or contains, an

25   affinity ligand that is capable of binding (preferably noncovalently, but covalent linkages are contemplated also) to a second, "capture" chemical moiety, such that a protein or peptide that naturally contains or is derivatized to include the affinity tag can be selected (or "captured") from a pool of proteins or peptides by contacting the pool with the capture moiety. The capture moiety is

30   preferably bound to a support surface, preferably a porous support surface, as a stationary phase. Examples of suitable supports include porous silica, porous titania, porous zirconia, porous organic polymers, porous polysaccharides, or any of these supports in non-porous form.

Preferably the interactions between the affinity tag and the capture moiety are specific and reversible (e.g., noncovalent binding or hydrolyzable covalent linkage), but they can, if desired, initially be, or subsequently be made, irreversible (e.g., a nonhydrolyzable covalent linkage between the affinity tag

5   and the capture moiety). It is important to understand that the invention is not limited to the use of any particular affinity ligand.

Examples of endogenous affinity ligands include naturally occurring amino acids such as cysteine (selected with, for example, an acylating reagent) and histidine, as well as carbohydrate and phosphate moieties. A portion of the

10   protein or peptide amino acid sequence that defines an antigen can also serve as an endogenous affinity ligand, which is particularly useful if the endogenous amino acid sequence is common to more than one protein in the original mixture. In that case, a polyclonal or monoclonal antibody that selects for families of polypeptides that contain the endogenous antigenic sequence can be

15   used as the capture moiety. An antigen is a substance that reacts with products of an immune response stimulated by a specific immunogen, including antibodies and/or T lymphocytes. As is known in the art, an antibody molecule or a T lymphocyte may bind to various substances, for example, sugars, lipids, intermediary metabolites, autocoids, hormones, complex carbohydrates,

20   phospholipids, nucleic acids, and proteins. As used herein, the term "antigen" means any substance present in a peptide that may be captured by binding to an antibody, a T lymphocyte, the binding portion of an antibody or the binding portion of T lymphocyte.

A non-endogenous (i.e., exogenous) affinity tag can be added to a

25   protein or peptide by, for example, first covalently linking the affinity ligand to a derivatizing agent to form an affinity tag, then using the affinity tag to derivatize at least one functional group on the protein or peptide. Alternatively, the protein or peptide can be first derivatized with the derivatizing agent, then the affinity ligand can be covalently linked to the derivatized protein or peptide

30   at a site on the derivatizing agent.

An example of an affinity ligand that can be covalently linked to a protein or peptide by way chemical or enzymatic derivatization is a peptide, preferably a peptide antigen or polyhistidine. A peptide antigen can itself be

39

derivatized with, for example, a 2,4-dinitrophenyl or fluorescein moiety, which renders the peptide more antigenic. A peptide antigen can be conveniently captured by an immunosorbant that contains a bound monoclonal or polyclonal antibody specific for the peptide antigen. A polyhistidine tag, on the other

5 hand, is typically captured by an IMAC column containing a metal chelating agent loaded with nickel or copper.

Biotin, preferably ethylenediamine terminated biotin, which can be captured by the natural receptor avidin, represents another affinity ligand. Other natural receptors can also be used as capture moieties in embodiments

10 wherein their ligands serve as affinity ligands. Other affinity ligands include dinitrophenol (which is typically captured using an antibody or a molecularly imprinted polymer), short oligonucleotides, and polypeptide nucleic acids (PNA) (which are typically captured by nucleic acid hybridization). Molecularly imprinted polymers can also be used to capture.

15 The affinity ligand is typically linked to a chemical moiety that is capable of derivatizing a selected functional group on a peptide or protein, to form an affinity tag. An affinity ligand can, for example, be covalently linked to maleimide (a protein or peptide derivatizing agent) to yield an affinity tag, which is then used to derivatize the free sulfhydryl groups in cysteine, as further

20 described below.

Selecting cysteine-containing peptides

As noted above, is a common strategy to alkylate the sulfhydryl groups in a protein before proteolysis. Alkylation is generally based on two kinds of

25 reactions. One is to alkylate with a reagent such as iodoacetic acid (IAA) or iodoacetamide (IAM). The other is to react with vinyl pyridine, maleic acid, or N-ethylmaleimide (NEM). This second derivatization method is based on the propensity of -SH groups to add to the C=C double bond in a conjugated system. Alkylating agents linked to an affinity ligand double as affinity tags

30 and can be used to select cysteine containing peptides after, or concomitant with, alkylation. For example, affinity-tagged iodoacetic acid is a convenient selection for cysteine.

Optionally, the protein is reduced prior to alkylation to convert all the
disulfides (cystines) into sulfhydryls (cysteines) prior to derivatization.
Alkylation can be performed either prior to reduction (permitting the capture of
only those fragments in which the cysteine is free in the native protein) or after
5   reduction (permitting capture of the larger group containing all cysteine-
containing peptides, include those that are in the oxidized cystine form in the
native protein).

Preparation of an affinity tagged N-ethylmaleimide may be achieved by
the addition of a primary amine-containing affinity tag to maleic anhydride.
10   The actual affinity tag may be chosen from among a number of species ranging
from peptide antigens, polyhistidine, biotin, dinitrophenol, or polypeptide
nucleic acids (PNA). Peptide and dinitrophenol tags are typically selected with
an antibody whereas the biotin tag is selected with avidin. When the affinity tag
includes as the affinity ligand a peptide, and when proteolysis of the protein
15   mixture is accomplished after derivatization using trypsin or lys-C, the peptide
affinity ligand preferably does not contain lysine or arginine, so as to prevent
the affinity ligand from also being cleaved during proteolysis. Biotin is a
preferred affinity ligand because it is selected with very high affinity and can be
captured with readily available avidin/streptavidin columns or magnetic beads.
20   As noted above, polyhistidine tags are selected in an immobilized metal affinity
chromatography (IMAC) capture step. This selection route has the advantage
that the columns are much less expensive, they are of high capacity, and
analytes are easily desorbed.

Alternatively, cysteine-containing peptides or proteins can be captured
25   directly during alkylation without incorporating an affinity ligand into the
alkylating agent. An alkylating agent is immobilized on a suitable substrate,
and the protein or peptide mixture is contacted with the immobilized alkylating
agent to select cysteine-containing peptides or proteins. If proteins are selected,
proteolysis can be conveniently carried out on the immobilized proteins to yield
30   immobilized cysteine-containing peptides. Selected peptides or proteins are
then released from the substrate and subjected to further processing in
accordance with the method of the invention.

When alkylation is done in solution, excess affinity tagged alkylating agent is removed prior to selection with an immobilized capture moiety. Failure to do so will severely reduce the capacity of the capture sorbent. This is because the tagged alkylating agent is used in great excess and the affinity

5   sorbent cannot discriminate between excess reagent and tagged peptides. This problem is readily circumvented by using a small size exclusion column to separate alkylated proteins from excess reagent prior to affinity selection. The whole process can be automated (as further described below) by using a multidimensional chromatography system with, for example, a size exclusion

10   column, an immobilized trypsin column, an affinity selector column, and a reversed phase column. After size discrimination the protein is valved through the trypsin column and the peptides in the effluent passed directly to the affinity column for selection. After capture and concentration on the affinity column, tagged peptides are desorbed from the affinity column and transferred to the

15   reversed phase column where they were again captured and concentrated. Finally, the peptides are eluted with a volatile mobile phase and fractions collected for mass spectral analysis. Automation in this manner has been found to work well.

20   *Selecting tyrosine-containing peptides*

Like cysteine, tyrosine is an amino acid that is present in proteins in limited abundance. It is known that diazonium salts add to the aromatic ring of tyrosine *ortho* to the hydroxyl groups; this fact has been widely exploited in the immobilization of proteins through tyrosine. Accordingly, tyrosine-containing

25   peptides or proteins can be affinity-selected by derivatizing them with a diazonium salt that has been coupled at its carboxyl group to a primary amine on an affinity ligand, for example through the α-amino group on a peptide tag as described above. Alternatively, that diazonium salt can be immobilized on a suitable substrate, and the protein or peptide mixture is contacted with the

30   immobilized diazonium salt to select tyrosine-containing peptides or proteins. If proteins are selected, proteolysis can be conveniently carried out on the immobilized proteins to yield immobilized tyrosine-containing peptides.

Selected peptides or proteins are then released from the substrate and subjected
to further processing in accordance with the method of the invention.

5   *Selecting tryptophan-containing peptides*

Tryptophan is present in most mammalian proteins at a level of <3%.
This means that the average protein will yield only a few tryptophan containing
peptides. Selective derivatization of tryptophan has been achieved with 2,4-
dinitrophenylsulfenyl chloride at pH 5.0 (M. Wilcheck et al., Biochem.
10   Biophys. Acta 178:1-7 (1972)). Using an antibody directed against 2,4-
dinitrophenol, an immunosorbant was prepared to select peptides with this label.
The advantage of tryptophan selection is that the number of peptides will
generally be small.

15   *Selecting histidine-containing peptides.*

In view of the higher frequency of histidine in proteins, it would seem at
first that far too many peptides would be selected to be useful. The great
strength of the procedure outlined below is that it selects on the basis of the
number of histidines, not just the presence of histidine. Immobilized metal
20   affinity chromatography (IMAC) columns loaded with copper easily produce
ten or more peaks. The fact that a few other amino acids are weakly selected is
not a problem, and the specificity of histidine selection can, if desired, be
greatly improved by acetylation of primary amino groups. Fractions from the
IMAC column are transferred to an RPC-MALDI/MS system for analysis. The
25   number of peptides that can potentially be analyzed jumps to 100,000-300,000
in the IMAC approach.

An automated IMAC-RPC-MALDI/MS system essentially identical to
that used for cysteine selection has been assembled. The only difference is in
30   substituting an IMAC column for the affinity sorbent and changes in the elution
protocol. Gradient elution in these systems is most easily achieved by applying
step gradients to the affinity column. After reduction, alkylation, and digestion,
the peptide mixture is captured on the IMAC column loaded with copper.

Peptides are isocratically eluted from the IMAC using imidazole or a change in
pH, and directly transferred to the RPC column where they are concentrated at
the head of the column. The IMAC is then taken off line, the solvent lines of
the instrument purged at 10 ml/minute for a few seconds with RPC solvent A,

5   and then the RPC column is gradient eluted and column fractions collected for
MALDI-MS. When this is done, the RPC column is recycled with the next
solvent for step elution of the IMAC column, the IMAC column is then brought
back on line, and the second set of peptides is isocratically eluted from the
IMAC column and transferred to the RPC column where they are readsorbed.

10   The IMAC column is again taken off-line, the system purged, and the second
set of peptides is eluted from the RPC column. This process is repeated until
the IMAC column has been eluted. Again, everything leading up to MALDI-
MS is automated.

15   *Selecting post-translationally modified proteins.*

Post-translational modification plays an important role in regulation.
For this reason, it is necessary to have methods that detect specific post-
translational modifications. Advantageously, the method of the invention can
distinguish among proteins having a single signature peptide where speciation

20   occurs by post-translational modification, if the affinity ligand is associated
with, or constitutes, the post-translational moiety (e.g., sugar residue or
phosphate). Among the more important post-translational modifications are i)
the phosphorylation of tyrosine, serine, or threonine; ii) N-glycosylation; and
iii) O-glycosylation.

25

Selecting phosphoproteins

In the case of phosphorylated proteins, such as those containing
phosphotyrosine and phosphoserine, selection can achieved with monoclonal
antibodies that target specific phosphorylated amino acids. For example,

30   immunosorbant columns loaded with a tyrosine phosphate specific monoclonal
antibody are commercially available. Preferably, all proteins in a sample are
digested, then the immunosorbant is used to select only the tyrosine phosphate

containing peptides. As in other selection schemes, these peptides can separated by reversed phase chromatography and subjected to MALDI.

Alternatively, selection of phosphopeptides can be achieved using IMAC columns loaded with gallium (M. Posewitz et al., Anal. Chem.

5 71(14):2883-2992 (1999)). Phosphopeptides can also be selected using anion exchange chromatography, preferably on a cationic support surface, at acidic pH.

In addition, because zirconate sorbents have high affinity for phosphate containing compounds (C. Dunlap et al., J. Chromatogr. A 746:199-210

10 (1996)), zirconia-containing chromatography is expected to be suitable for the purification of phosphoproteins and phosphopeptides. Zirconate clad silica sorbents can be prepared by applying zirconyl chloride dissolved in 2,4-pentadione to 500 angstrom pore diameter silica and then heat treating the support at 400EC. Another alternative is the porous zirconate support recently

15 described by Peter Carr (C. Dunlap et al., J. Chromatogr. A 746:199-210 (1996)). Phosphopeptides are eluted using a phosphate buffer gradient. In many respects, this strategy is the same as that of the IMAC columns.

Selecting O-linked oligosaccharide containing peptides

20 Glycopeptides can be selected using lectins. For example, lectin from *Bandeiraea simplicifolia* (BS-II) binds readily to proteins containing N-acetylglucosamine. This lectin is immobilized on a silica support and used to affinity select O-glycosylated proteins, such transcription factors, containing N-acetylglucosamine and the glycopeptides resulting from proteolysis. The

25 protocol is essentially identical to the other affinity selection methods described above. Following reduction and alkylation, low molecular weight reagents are separated from proteins. The proteins are then tryptic digested, the glycopeptides selected on the affinity column, and then the glycopeptides resolved by RPC. In the case of some transcription factors, glycosylation is

30 homogeneous and MALDI-MS of the intact glycopeptide is unambiguous. That is not the case with the more complex O-linked glycopeptides obtained from many other systems. Heterogeneity of glycosylation at a particular serine will produce a complex mass spectrum that is difficult to interpret. Enzymatic

deglycosylation of peptides subsequent to affinity selection is indicated in these cases. Deglycosylation can also be achieved chemically with strong base and is followed by size exclusion chromatography to separate the peptides from the cleaved oligosaccharides.

5   It is important to note that O-linked and N-linked glycopeptides are easily differentiated by selective cleavage of serine linked oligosaccharides (E. Roquemore et al., Meth. Enzymol. 230:443-460 (1994)). There are multiple ways to chemically differentiate between these two classes of glycopeptides. For example, basic conditions in which the hemiacetal linkage to serine is

10   readily cleaved can be utilized. In the process, serine is dehydrated to form an α,β unsaturated system (C=C-C=O). The C=C bond of this system may be either reduced with $NaBH_4$ or alkylated with a tagged thiol for further affinity selection. This would allow O-linked glycopeptides to be selected in the presence of N-linked glycopeptides. The same result could be achieved with

15   enzymatic digestion.

Selecting N-linked oligosaccharide-containing peptides

          As with O-linked oligosaccharide-containing peptides, lectins can be used to affinity select N-linked glycopeptides following reductive alkylation

20   and proteolysis. To avoid selecting O-linked glycopeptides, the peptide mixture is subjected to conditions that cause selective cleavage O-linked oligosaccharides prior to affinity selection using the lectin. Preferably O-linked deglycosylation is achieved using a base treatment after reductive alkylation, followed by size exclusion chromatography to separate the peptides from the

25   cleaved oligosaccharides. To address the potential problem of heterogeneity of glycosylation, and N-linked glycopeptides are deglycosylated after selection. Automation can be achieved with immobilized enzymes, but long residence times in the enzyme columns are needed for the three enzymatic hydrolysis steps.

30

Identification of signature peptides and their parent proteins

After peptides of interest are detected using mass spectrometry, for example by identifying and determining the mass of those peptides that show a different abundance ratio, it is often desired to identify the protein from which a

5 peptide originated. In most instances this can be accomplished using a standard protocol that involves scanning either protein or DNA databases for amino acid sequences that would correspond to the proteolytic fragments generated experimentally, matching the mass of all possible fragments against the experimental data (F. Hsieh et al., Anal. Chem. 70:1847-1852 (1998); D. Reiber

10 et all, Anal. Chem 70:673-683 (1998)). When a DNA database is used as a reference database, open reading frames are translated and the resulting putative proteins are cleaved computationally to generate the reference fragments, using the same cleavage method that was used experimentally. Likewise, when a protein database is used, proteolytic cleavage is also performed computationally

15 to generate the reference fragments. In addition, masses of the reference peptide fragments are adjusted as necessary to reflect derivatizations equivalent to those made to the experimental peptides, for example to include the exogenous affinity tag.

The presence of signature peptides in the sample is detected by

20 comparing the masses of the experimentally generated peptides with the masses of signature peptides derived from putative proteolytic cleavage of the set of reference proteins obtained from the database. Software and databases suited to this purpose are readily available either through commercial mass spectrometer software and the Internet. Optionally, the peptide databases can be preselected

25 or reduced in complexity by removing peptides that do not contain the amino acid(s) upon which affinity selection is based.

There will, of course, be instances where peptides cannot be identified from databases or when multiple peptides in the database have the same mass. One approach to this problem is to sequence the peptide in the mass

30 spectrometer by collision induced dissociation. Ideally this is done with a MALDI-MS/MS or ESI-MS/MS instrument. Another way to proceed is to isolate peptides and sequence them by a conventional method. Because the signature peptide strategy is based on chromatographic separation methods, it is

47

generally relatively easy to purify peptides for amino acid sequencing if sufficient material is available. For example, conventional PTH-based sequencing or carboxypeptidase based C-terminal sequencing described for MALDI-MS several years ago (D. Patterson et al., Anal. Chem. 67:3971-3978

5   (1995)). In cases where 6-10 amino acids can be sequenced from the C-terminus of a peptide, it is often possible to synthesize DNA probes that would allow selective amplification of the cDNA complement along with DNA sequencing to arrive at the structure of the protein.

10   EXAMPLES

The present invention is illustrated by the following examples. It is to be understood that the particular examples, materials, amounts, and procedures are to be interpreted broadly in accordance with the scope and spirit of the

15   invention as set forth herein.

Example 1.

Fractionation of Isotopically Labeled Peptides in Quantitative Proteomics

20   This study examines the extent to which isotopic forms of peptides having the same amino acid sequence are resolved by reversed-phase chromatography and assesses the degree to which resolution of these isotopically different forms of a peptide impact quantification.

MATERIALS AND METHODS

25   *Materials.* Human angiotensin I (Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu-OH), HPLC-grade acetonitrile (ACN), N-hydroxysuccinimide, N-acetoxysuccinimide, monobasic sodium phosphate, dibasic sodium phosphate, acetic-d_6-anhydride, succinic anhydride, and succinic-d_4-anhydride were purchased from Sigma-Aldrich Co. (St. Louis, MO). The PepMap C18 4.6 i.d.

30   (internal diameter) x 250 mm was purchased from Applied Biosystems (Framingham, MA). Double-deionized water (ddI H_2O) was produced by a Milli-Q Gradient A10 System from Millipore Co. (Bedford, MA). All reagents were used directly without further purification.

48

Synthesis of N-Acetoxy-d₃-succinimide. A solution of 4.0 g (34.8 mmol) of *N*-hydroxysuccinimide in 11.4 g (105 mmol) of acetic-d_6-anhydride was stirred at room temperature. White crystals began to deposit in 10 minutes. After 15 hours the solution was filtered. The crystals were washed with hexane

5   and then dried in a vacuum. The product yield was 5.43 g (100%), mp 133-134 °C (Ji et al., J. Chromatogr. B 2000, 745, 197–210).

Acetylation of peptides. A 3-fold molar excess of *N*-acetoxysuccinimide and *N*-acetoxy-d_3-succinimide was added individually to equal aliquots of 1 mg/mL peptide solution in phosphate buffer at pH 7.5. After 4-5 hours stirring

10 at room temperature, equal aliquots of the two samples were mixed (Ji et al., J. Chromatogr. B 2000, 745, 197–210).

Isotopic labeling of peptides by succinic anhydride and succinic-d_4-anhydride. A 50-fold molar excess of succinic anhydride and succinic-d_4-anhydride was added individually to experimental and control samples in

15 phosphate buffer, pH 7-8. Labeling reagent was added in small aliquots over the course of the first hour, and the reaction was allowed to proceed for another 2 hours. N-Hydroxylamine was then added in excess, and the pH was adjusted to 11-12. Incubation with hydroxylamine was allowed to proceed for 10 minutes. The function of the hydroxylamine reaction was to hydrolyze esters

20 that might have been formed during the acylation reaction.

Reversed-phase elution of isotopically labeled peptides. Isotopically labeled peptide mixtures were separated by gradient elution from a 4.6 i.d. x 250 mm PepMap C18 silica column on an Integral Micro-Analytical Workstation (Applied Biosystems, Framingham, MA). The PepMap C18

25 column was equilibrated using 95% mobile phase A (0.1% TFA in ddI H₂O) and 5% mobile phase B (95% ACN/0.1% TFA in ddI H₂O) at a flow rate of 1.00 mL/min for 3 column volumes (CV). A 100-μL portion of the isotopically labeled peptides was injected automatically and gradient elution of the analyte was achieved using 100% mobile phase A to 100% mobile phase B over 60

30 minutes at a flow rate of 1.00 mL/min. The gradient was then held at 100% mobile phase B for an additional 5 minutes. Throughout the analysis, an on-line UV detector set at 214 nm was used to monitor isotopic separation of the

peptide mixture. The peptides were simultaneously monitored by ESI-MS directly coupled to the LC.

ESI-MS analysis. Mass spectral analyses were performed using a Mariner Biospectrometry Workstation (Applied Biosystems, Framingham, MA)

5   equipped with an atmospheric pressure ESI source. All spectra were obtained in the positive ion mode at a sampling rate of 1 spectrum every 3 seconds. During the coupled ESI-MS acquisition, masses were scanned from m/z 500 to 3000 at 140 V nozzle potential. The instrument was tuned and mass calibrated by manual injection of Mariner calibration mixture (Applied Biosystems,

10   Framingham, MA) at 3.0 μL/min. Extracted ion chromatograms were used to reconstruct elution profiles of deuterated and nondeuterated peptides separately.

RESULTS AND DISCUSSION

The goals of this study were to (1) assess the extent to which resolution of internal standards and analytes would impact quantification, (2) develop methods to quantify partial resolution of components in nonsymmetrical

5   chromatographic peaks, and (3) examine experimentally the degree to which resolution of several isotopically labeled species occurred.

Theoretical assessment of isotopic fractionation. Simulations were used to examine the influence of isotopic fractionation on abundance ratio quantification. In the simulation below, deuterated and nondeuterated peptide

10   chromatographic peaks were assumed to have a Gaussian shape. Deuterated compounds usually elute earlier than their nondeuterated counterparts in reversed-phase liquid chromatography. Assuming Gaussian peak shape, the elution profile of the deuterated peptide peak can be described by the equation

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{t^2}{\sigma^2}} \tag{1}$$

15   but that of the nondeuterated peptide peak would be

$$f(t - \Delta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(t-\Delta)^2}{\sigma^2}} \tag{2}$$

where t is time relative to the center of the deuterated peptide peak, f(t) is peak intensity of the deuterated peptide at time t, f(t-Δ) is peak intensity of the nondeuterated peptide at time t, Δ is the separation between the deuterated and

20   nondeuterated peptide peaks (which was positive in all cases studied), σ is standard deviation and is equal to $W_{1/2}/2.355$ for Gaussian peaks, and $W_{1/2}$ is the full peak width at half-maximum of a Gaussian peak. When resolution (R) between these two compounds is expressed as

$$R = \frac{\Delta}{W_{1/2}} = \frac{\Delta}{2.355\sigma} \tag{3}$$

25   the observed abundance ratio (ratio_{obs}, nondeuterated peptide vs deuterated peptide, i.e., H/D) at any point in the elution profile of the two analytes will be given by the equation

$$ratio_{obs} = ratio_{true} \frac{f(t - \Delta)}{f(t)} = ratio_{true} e^{\frac{\Delta}{2\sigma^2}(2t-\Delta)} = ratio_{true} e^{2.355R\frac{t}{\sigma}-2.773R^2} \tag{4}$$

where ratio$_{true}$ is the abundance ratio in the sample before chromatography.
Following rearrangement and substitution of terms, equation 4 becomes

$$\ln(ratio_{obs}) = \frac{2.355R}{\sigma}t + (\ln ratio_{true} - 2.773R^2) \tag{5}$$

and

$$R = slope\frac{\sigma}{2.355ratio_{true}} = \frac{(slope)(W_{1/2})}{5.546} \tag{6}$$

Eqs. 5 and 6 suggest that a plot of ln(ratio$_{obs}$) vs. t will be linear and the
slope could be used to calculate R when peak asymmetry is small. Plots of
experimental data using this model will be referred to as a *Gaussian model log
plot*.

Using a Taylor series expansion when R and $|t|$ are small,

$$ratio_{obs} \approx ratio_{true}(1 + 2.355R\frac{t}{\sigma} - 2.773R^2)$$
$$= \frac{2.355Rratio_{true}}{\sigma}t + (ratio_{true} - 2.773ratio_{true}R^2) \tag{7}$$

and

$$R = slope\frac{\sigma}{2.355ratio_{true}} = \frac{(slope)(W_{1/2})}{5.546ratio_{true}} \tag{8}$$

According to Eqs. 7 and 8, a plot of ratio$_{obs}$ vs t will be linear at the
center of the peak ($\pm1\sigma$), and the slope of the straight line obtained could be
used to calculate R when R and peak asymmetry are small. Plots derived using
this model will be designated as a *Gaussian model linear plot*.

Simulation of the elution of two substances from a chromatography
column as Gaussian curves is seen in Figure 1. Ratio$_{obs}$ is constant with time
and always equals ratio$_{true}$ when R = 0 (Figure 1A). Assuming that R = 0.025
and 0.5, the deuterated peptide elutes first, and the full peak width at half-
maximum ($W_{1/2}$) of the peak is 60 seconds, the deuterated peptide would elute
1.5 and 30 seconds faster than the nondeuterated peptide and ratio$_{obs}$ would vary
continuously across the peak (Figure 1B, C). This leads to a substantial error in
determining the abundance ratio in a sample when quantification measurements
are made at the peak extremes. It is seen in Table 1 that a measurement made at
$\pm1\sigma$ would deviate by −83 and +500% from the abundance ratio in the original

sample (ratio$_{true}$) when R = 0.5. The error would be even greater at larger values of σ.

Table 1. Examples of Systematic Quantification Errors Resulting from

Fractionation of Isotopically Labeled Peptides

measurement point	R	ratio$_{obs}$	ratio$_{true}$	error%
t = -1σ	0.025	0.94	1.0	-6%
	0.5	0.17	1.0	-83%
t = +1σ	0.025	1.06	1.0	+6%
	0.5	6.0	1.0	+500%
t = 0 (center of peak)	0.025	1.0	1.0	0
	0.5	1.0	1.0	0
integration of first half of the peak	0.025	0.96	1.0	-4%
	0.5	0.39	1.0	-61%
integration of second half of the peak	0.025	1.04	1.0	+4%
	0.5	2.6	1.0	+160%
integration across the whole peak	0.025	1.0	1.0	0
	0.5	1.0	1.0	0

There is the possibility that only a portion of a peak will be collected and sent into subsequent dimensions of analysis, because multidimensional separations are used in proteomics. Assuming that only half of the peak was sent into a second chromatographic dimension in the case described in Figure 1C, the leading edge of the peak would have 44% more of the deuterated species and the trailing edge 44% less than that found in the original sample (Table 1). The opposite is true for the nondeuterated species. Table 1 also shows that accurate ratio measurements may be made either at the center of the peak or by integrating the whole peak. Unfortunately, these options are not always available.

There are many cases in which partial resolution of isotopically labeled peptides might be a problem. The multidimensional chromatography case has been noted above. Similar problems would arise with the collection of fractions

53

across a chromatographic peak (Figure 1B, C) for analysis by matrix assisted laser desorption ionization mass spectrometry (MALDI-MS). The abundance ratio in fractions collected at large σ values would deviate substantially from the "true" ratio in the sample. Another potential problem is that matrix components eluting with the analyte at leading and trailing edges of peaks can be very different. This introduces the potential for differences in ionization efficiency across a peak. In this case, integration of the whole peak or taking measurements at the center of the peak to achieve accurate ratio measurement becomes invalid.

Table 2 presents another view of the problem. Assuming R = 0.5 and the ratio measurement is made at $\pm 1\sigma$ from the center of the peak, it is seen in Table 2 that there is little difference between the ratio$_{obs}$ value of 0.18 at $+1\sigma$ from a sample in which ratio$_{true}$ is 0.1 and the ratio$_{obs}$ value of 0.17 at -1σ from a sample in which ratio$_{true}$ is 1.0. A similar problem is seen in differentiating between the ratio$_{obs}$ values of 1.0 and 10 at their respective time of $+1\sigma$ and -1σ. When R is larger or ratio measurements are made further away from the center of the peak, the problem is even more serious!

Table 2. Systematic Errors Caused by Fractionation of the Deuterated and Nondeuterated Peptides in Reversed-Phase Chromatography, R = 0.5

ratio$_{true}$	ratio$_{obs}$	
	-1σ	$+1\sigma$
1:10 (down-regulated)	0.017	0.18
1:1 (no change)	0.17	6.0
10:1 (up-regulated)	5.8	60

Calculation of R. It has been seen in Figure 1 that even small differences in R can have a substantial impact on abundance ratio determinations. Unfortunately, chromatography peaks are frequently not Gaussian, and Eqs. 6 and 8 will not predict R with the accuracy needed to evaluate small differences in R. Although Gaussian models still provide good estimations, a more general method is needed to evaluate resolution (R). The Δ

54

tuning method described below allows evaluation of R with peaks of any shape. Assuming that deuterated and nondeuterated peptides have the same peak shape and plotting integrated peak area against time, ratio$_{obs}$ is redefined as the integrated peak area of the nonduterated peptide over that of deuterated peptide

5   (H/D) at time t in this model. When $\dot{R}$ = 0, ratio$_{obs}$ obviously does not vary with time (Figure 2A).

When the deuterated peptide is partially resolved from the nondeuterated peptide, ratio$_{obs}$ increases with elution time and eventually reaches a plateau at ratio$_{true}$ (Figure 2B). Assuming that the two elution curves are of the same

10   shape, they may be brought to overlap if the elution curve of the deuterated peptide is normalized (i.e., multiplied by ratio$_{true}$) and shifted in time until the two curves coincide. The shifting in time, or Δ *tuning*, required to cause overlap is given by the expression

$$\Delta = R * W_{1/2} \tag{9}$$

15   Retention times of the deuterated peptide (t_D) and nondeuterated peptide (t_H) can be measured directly from the extracted ion chromatograms produced by an electrospray ionization (ESI) mass spectrometer, and R can be calculated using the equation

$$R = \frac{t_H - t_D}{W_{1/2}} \tag{10}$$

20   when R is large.

Table 3 summarizes the options for calculating R in four different cases. It is seen that the Δ *tuning* method is the most general, applying to all four cases.

25

Table 3. Calculation of Resolution (R) in Different Situations

	R	peak asymmetry	methods apply
case I	small	small	Gaussian model linear plot; Gaussian model log plot; Δ tuning;
case II	small	large	Δ tuning;
case III	large	small	t_H-t_D; Gaussian model log plot; Δ tuning;
case IV	large	large	t_H-t_D; Δ tuning;

5   *Comparison of R values from three different labeling reagents.* R values
from three different labeling reagents, succinic anhydride (d_0 and d_4), N-
acetoxysuccinimide (d_0 and d_3), and an isotope-coded affinity tag (i.e., ICAT, d_0
and d_8) (Parker et al., 48th ASMS Conference on Mass Spectrometry and Allied
Topics, 2000) were determined using the methods in Table 3.

10   Angiotensin I samples were labeled with succinic anhydride and
succinic-d_4-anhydride and mixed. Peak asymmetry was small, and the plots of
$ratio_{obs}$ vs time and $ln(ratio_{obs})$ vs. time showed good linearity in Figure 3A.
Calculation of R based on the *Gaussian model linear plot* and *Gaussian model
log plot* both equal 0.088 (Eqs. 6 and 8). The *Δ tuning* method was applied in

15   Figure 3B, C. It was necessary to shift the elution curve of the deuterated
peptide 1.94 seconds to the right to make the two curves coincide. The R value
based on the *Δ tuning* method was 0.073, close to the R values obtained with the
Gaussian model linear plot and *Gaussian model log plot* (i.e., 0.088).

At low peptide concentration, the peak intensity plot was found to be

20   very noisy, and the integrated peak area plot greatly improved accuracy and
precision of the R measurement. In addition, because *Δ tuning* is based on only
one assumption that deuterated and nondeuterated peptides have the same peak
shape, the method will be generally applicable to the determination of R values.

56

Angiotensin 1 was also labeled with CH_3CO- and CD_3CO- groups. The sample was intentionally over-derivatized to produce an analyte with a single acetate group (Figure 4A) and another with two acetate groups (Figure 4B). R values calculated using the Δ *tuning* method were 0.023 and 0.11, respectively.

5 A resolution of 0.023 is roughly equivalent to that seen in Figure 1B. The large increase of R when six deuterium atoms were incorporated, as compared to three, could result from the fact that the second acetylation was on tyrosine. Tyrosine is hydrophobic and interacts heavily with the stationary phase in reversed-phase chromatography.

10 In the literature, ICAT d_0- and d_8-labeled peptides were shown to be separated (Parker et al., 48th ASMS Conference on Mass Spectrometry and Allied Topics, 2000). $Ratio_{obs}$ kept increasing across the peak. R was estimated to be 0.45 on the basis of both the t_H-t_D and *Gaussian model log plot* methods. It was assumed that the deuterated peptide and nondeuterated peptide reached

15 their peak maxima at spectrum numbers 390 and 400, respectively (Table 4). A resolution of 0.45 is roughly equivalent to that seen in Figure 1C.

Table 4. $Ratio_{obs}$ in Selected Spectra of ICAT-d_0 and d_8 Labeled Peptide (Adapted from Parker et al., 48th ASMS Conference on Mass Spectrometry and

20   Allied Topics, 2000).

spectrum #	d_0 mass	intensity	d_8 mass	intensity	ratio (H/D)
390	592.2802	5385	594.9623	9953	0.54
400	592.2808	10066	594.9619	6370	1.6
410	592.2802	3896	594.9616	1011	3.8

R values under different conditions for all of the labeling reagents studied are summarized in Table 5. It is obvious that R varies substantially with

25 different labeling reagents and the number of deuterium atoms incorporated.

Table 5. Comparison of R under Different Conditions*

	succinic anhydride	N-acetoxysuccinimide		ICAT
no. ^{2}H atoms incorporated	4	3	6	8
R	0.073	0.023	0.11	0.45

* Data from all three labeling agents were derived under slightly different chromatographic conditions. The extent to which this impacts resolution of isotopic isoforms is unknown.

Significance of fractionation of isotopically labeled peptides in quantitative proteomics. Suppression of ionization between peptides has been noted in ESI-MS when total peptide concentration is high, as when one peptide is eluting in a large background of another. This means that ionization efficiency can vary across a chromatographic peak. Similar problems are seen in MALDI-MS in which matrix peptides have a large impact on suppression. The whole chromatographic peak may not be in the fraction being examined, because samples for MALDI-MS are obtained by fraction collection. As the deuterated and nondeuterated peptides separate from each other, they are ionized at different times and in different matrixes. They effectively become external standards instead of internal standards. This has the potential to produce both significant systematic errors and much greater random errors. In addition, as the elution time becomes longer, the danger of overlapping with other peptides with similar mass increases.

There are other important reasons to minimize R. It is very difficult to do MS/MS analyses on every peptide as it elutes from a chromatography column, particularly when sample fractions contain 10-50 peptides. When the objective is to analyze only a small percentage of peptides that are up- or down-regulated, rapid quantification of changes in concentration based on a single mass spectrum would allow near real-time decisions of whether to execute an MS/MS analysis of a peptide. To accomplish this, ratio$_{true}$ must be determined accurately as early as possible so that the up- or down-regulated peptides can be

58

analyzed by MS/MS during elution of the rest of the peak. This is impossible when $ratio_{obs}$ varies with time. This would mean that the LC-MS analysis has to be done twice on each sample. The first time would be for abundance ratio analysis and the second, for MS/MS on selected peptides. Appropriate stable

5   isotope labeling reagents are desired to make R minimal so that $ratio_{obs}$ stays constant or varies insignificantly with time.

CONCLUSIONS

It has been demonstrated both theoretically and experimentally that the chromatographic fractionation of deuterated and nondeuterated peptides could

10   cause substantial systematic errors in quantitative proteomics.

The data were examined in four different ways. The most generally useful treatment of the data was the one in which extracted ion elution data was integrated and resolution was assessed by the time shift required to cause the curves from isotopically different forms of a peptide to coincide (i.e., Δ *tuning*).

15   Acetylated peptides showed the lowest degree of separation. Resolution of the deuterated and nondeuterated forms in this case was 0.023. In contrast, resolution of the deuterated and nondeuterated forms of the ICAT reagent were calculated to be 0.45. The deuterium content of the labeling reagent appears to play a major role. Deuterium content should be minimized to reduce isotopic

20   fractionation problems.

Example 2.

Minimizing Chromatographic Resolution Using $^{13}C/^{12}C$-coded Peptide Isoforms

25

This experiment was conducted to determine the magnitude of the isotope effect with a commercial deuterated coding agent and to assess the degree to which isotope effects can be minimized with a ^{13}C labeled derivatizing agent.

30

MATERIALS AND METHODS

Materials. Bovine serum albumin (BSA), HPLC grade acetonitrile (ACN), succinic anhydride, acetyl chloride, phosphorus pentoxide, dithiothreitol (DTT), iodoacetic acid (IAA), urea,

5   tris(hydroxymethy)aminomethane (Tris base), tris(hydroxymethyl)aminomethane hydrochloride (Tris acid), calcium chloride were purchased from Sigma-Aldrich (St. Louis, MO). Trifluoroacetic acid (Sequanal Grade) was obtained from Pierce (Rockford, IL). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Butanedioic

10   acid-$^{13}C_4$ was supplied by Isotec (Miamisburg, OH). The ICAT kit was purchased from Applied Biosystems (Framingham, MA). A C18 column (2.1 mm x 250 mm) was obtained from Vydac (Hesperia, CA). Double deionized water (ddI H_2O) was produced by a Milli-Q™ Gradient™ A10 System from Millipore (Bedford, MA).

15   *Proteolysis of BSA*. BSA (5 mg) was reduced in 1 mL of 0.2 mol/L Tris buffer (pH 8.5) containing 8 M urea and 10 mM DTT. After a 2 hour incubation at 37°C, iodoacetic acid was added to a final concentration of 20 mmol/L and incubated in darkness on ice for 2 more hours. After dilution with 0.2 mol/L Tris buffer to a final urea concentration of 2 M, modified sequencing

20   grade trypsin was added to the sample at a 50/1 BSA to trypsin mass ratio and the solution was incubated for 8 hours at 37°C. Digestion was stopped by freezing the mixture in liquid nitrogen for 10 minutes.

Synthesis of succinic anhydride-$^{13}C_4$. 2 g of butanedioic-$^{13}C_4$-acid and 5 mL of acetyl chloride were heated under reflux for 1.5 hours. After reaction the

25   clear solution was cooled to room temperature and kept at 0 °C overnight. The crystals formed upon cooling were further purified from dry diethyl ether and dried in *vacuo* over P_2O_5.

Derivatization of peptides. A fifty fold molar excess of succinic anhydride and succinic anhydride-$^{13}C_4$ were added individually to tryptic

30   peptides from experimental and control samples. The reaction was allowed to proceed for two hours at room temperature. Peptides were also derivatized with ICAT and ICAT-$^2H_8$ as described in the instructions of ICAT kit.

60

Reversed-phase chromatography of isotopically labeled peptides.
Isotopically labeled peptide mixtures were separated by gradient elution from a
Vydac C18 column (2.1 mm x 250 mm) on an Integral Micro-Analytical
Workstation (Applied Biosystems, Framingham, MA). The C18 column was

5   equilibrated using 100% mobile phase A (0.01% TFA in ddl H_2O) at a flow rate
of 250 µL/min for 2 column volumes (CV). Isotopically labeled peptide
mixtures (2 nmol) were injected and eluted at a flow rate of 250 µL/min in a
linear gradient ranging over 60 min from 100% mobile phase A to 60% mobile
phase B (95% ACN/0.01% TFA in ddl H_2O). At the end of this period a second

10   linear gradient was applied in 10 minutes from 60% B to 100% B at the same
flow rate. The gradient was then held at 100% mobile phase B for an additional
10 minutes. Throughout the analysis an on-line UV detector set at 214 nm was
used to monitor separation of the peptide mixtures. The peptides were
simultaneously monitored by ESI-MS by directing 10% of the flow into the

15   mass spectrometer.

      ESI-MS analysis. Mass spectral analyses were performed using a
QSTAR workstation (Applied Biosystems, Framingham, MA) equipped with an
Ionspray source. All spectra were obtained in the positive ion TOF mode at a
sampling rate of one spectrum every two seconds. During LC-MS data

20   acquisition, masses were scanned from *m/z* 300 to 1800.

RESULTS AND DISCUSSION

      Assessing the impact of heavy isotopes on chromatographic behavior
requires quantification of small changes in resolution (R). Moreover, peaks are

25   not always Gaussian and it is important to deal with peak asymmetry in
computing resolution. A "*Δ Tuning*" method is capable of determining
resolution in the case of both symmetrical and asymmetric peaks. It was chosen
for the studies outlined below for this reason. This method is based on shifting
integrated plots of isoform elution profiles until they overlap. Resolution (R)

30   from the shifting in time, or "*Δ tuning*" required to cause overlap is give by the
expression

$$R = \frac{\Delta}{W_{1/2}} \qquad\qquad (11)$$

61

where Δ is the time shift required to cause the heavy isotope labeled analyte and the non-isotope labeled analyte elution curves to overlap and $W_{1/2}$ is the average of full peak width at half maximum (FWHM) derived from extracted ion chromatograms of the isoforms. Even the peak shapes of the isotopically

5   labeled peptides are sometime very different when resolution is large. In this case "Δ" simply equals the differences in retention times and "$W_{1/2}$" still equals the average peak widths.

The possibility that isotope effects could vary between peptides must also be considered. For example, peptides with multiple derivatizable

10   functional groups would be expected to show larger isotope effects. The location of the derivatizing agent relative to hydrophobic groups in the peptide could also play a role. Rather than select specific peptides to test these hypotheses, a tryptic digest of bovine serum albumin (BSA) was chosen for study.

15   *Isotope effects with ICAT.* The isotope coded affinity tag (ICAT) reagent is offered commercially for differentially labeling peptides. This reagent is a sulfhydryl directed alkylating agent composed of iodoacetate attached to biotin through a coupling arm. Isotopic isoforms of this reagent are differentially labeled in the coupling arm (Figure 5A). During the normal course of reducing

20   and alkylating proteins in preparation for proteolysis, the ICAT reagent is used to alkylate cysteine residues in proteins. Subsequent to differential labeling of control and experimental samples with the $^2H_0$ and $^2H_8$ versions of the ICAT reagent respectively, the samples are mixed and digested with a proteolytic enzyme. Biotinylated, cysteine-containing peptides were selected from digests

25   and the relative concentration of the isotopic isoforms of peptides quantified by mass spectrometry. Approximately 10-20% of all peptides derived from the proteome of eukaryotes contain one or more cysteine residues. This means that the ICAT reagent can only be used to quantify changes in a proteome that involve cysteine-containing peptides.

30   An example of isotopic fractionation of ICAT-$^2H_0$ and $^2H_8$ labeled peptides is seen in the case of QNCDQFEK (SEQ ID NO:1). Derivatization was at the single cysteine residue in the peptide. The $^2H_8$ labeled peptide eluted approximately 28 seconds earlier than the non-deuterated peptide, causing an

enormous variation in abundance ratio across the elution profile of the isoforms (Figure 6A-D). Resolution of the isoforms was calculated to be 0.74. Even the chromatographic peak shapes of the deuterated and non-deuterated peptides can differ (Figure 6A). It is readily seen that the only accurate way to quantify the

5   abundance ratio in the initial sample is through a comparison of the integrated extracted ion chromatograms of the isoforms.

Resolution of tryptic peptides from a digest of bovine serum albumin is seen in Figure 8. Twenty percent of the peptides examined in this limited study showed a resolution of greater than 0.5. Of this number, three fourth were

10   derivatized with two molecules of ICAT reagent. In fact, half the peptides derivatized with two molecules of ICAT reagent showed a resolution of greater than 0.5. In contrast, only 23% of singly labeled peptides had a resolution greater than 0.4. It is also seen in Figure 8 that isotope effects tend to become smaller with increasing molecular weight of the peptide, particularly with those

15   that are singly derivatized.

Resolution of peptide isoforms is an important concern for a number of reasons. One is the matter of ionization efficiency across the elution profile of the isoforms. It is probable that in a complex mixture that peptides eluting in the leading and tailing parts of a peak are different. If sample matrix

20   components diminish ionization efficiency more in one part of an analyte peak than another does, there is a problem in ESI-MS. Integrated peak areas will not be accurate in cases where the ionization efficiency of a compound varies during peak elution. Moreover, ionization efficiency could vary between the isoforms when they are partially resolved for the same reason.

25   A similar problem occurs with MALDI-MS. Large numbers of peaks will overlap in the reversed phase elution profile of complex tryptic digests and there will be no distinct peaks. This means that fraction collection for MALDI-MS from chromatography columns will be time based and there is a high probability components will inadvertently be split into multiple fractions.

30   When isotopic isoforms are differentially resolved in these fractions they will differ substantially in abundance ratio as the simulation in Figures 6E-G shows. The problem is in knowing the portion of the analyte present in each fraction that is needed to calculate the true abundance ratio in the initial sample.

Still another problem is that it is necessary to wait until both isoforms
elute before extracted ion chromatograms can be constructed and peak areas can
be calculated to determine abundance ratio. When the objective is to identify
only those components in the mixture that have been up or down-regulated and

5   there is a need for MS/MS data to identify a peptide that has changed in
concentration, the isoforms will have eluted before it is known that MS/MS
sequence data is needed. This means that either MS/MS data must be acquired
on all components as they elute, which is difficult to do, or the sample must be
run a second time to acquire the necessary MS/MS data, which costs more time

10   and sample.

Isotope effects with ^{13}C succinate. Succinic anhydride (Figure 5A)
derivatizes primary amine groups in peptides at their N-termini and in lysine
residues. This means that each peptide produced during proteolysis can be
derivatized. The exception would be rare peptides 1) that do not contain a

15   lysine residue and 2) are derived from the amino terminus of an amino
terminally blocked protein. Differential labeling of tryptic digests from control
and experimental samples with $^{13}C_0$ and $^{13}C_4$-succinic anhydride respectively,
produces peptide isoforms that vary by 4 and 8 atomic mass units (amu).
Peptides with a C-terminal arginine appear in mass spectra as a double cluster

20   of ions separated by 4 amu whereas C-terminal lysine-containing peptides are
separated by 8 amu. In the case of the peptide QNCDQFEK (SEQ ID NO:1),
two moles of succinate were added when it was derivatized with succinic
anhydride. The reversed-phase chromatography elution profile of the $^{13}C_0$ and
$^{13}C_4$-succinate labeled isoforms of QNCDQFEK (SEQ ID NO:1) is seen in

25   Figure 7A. Resolution of the isoforms was calculated to be -0.0043. This is
within the measurement error of the method. The fact that there is no detectable
resolution of the isoforms means that the abundance ratio will be constant
across the elution profile (Figures 7A-D). Resolution of all the peptide isoforms
from a tryptic digest of BSA falls between +0.01 and –0.01 (Figure 9), again

30   within experimental error.

It is important to know that there was no evidence of peptide
fractionation based on the content of natural heavy isotopes in peptides. In view
of the fact that all peptides contain small amounts of ^{13}C and ^2H from the

64

occurrence of these isotopes in nature, it is interesting that they were not observed to resolve in reversed phase chromatography. This is probably because the isotope peaks of unlabeled peptides are due mainly to ^{13}C. The natural abundance of ^{13}C is 1.10% whereas that of ^{2}H is 0.015%. Although

5 peptides have roughly twice as many hydrogen as carbon atoms, few peptides would contain more than one or two deuterium atoms. The lack of isotope effects in the fractionation of ^{13}C labeled peptides and the corresponding natural monoisotopic peptides means they could be added together to calculate the ratio between experimental and control samples to improve the precision of

10 abundance ratio measurement.

A direct comparison of peptide resolution caused by succinic anhydride-$^{13}C_4$ and ICAT-$^{2}H_8$ in the same group of peptides is seen in Table 6. ICAT-$^{2}H_8$ caused significant resolution between the isotopically labeled peptides, while succinic anhydride-$^{13}C_4$ did not. Clearly, coding labeling agents with ^{13}C is

15 vastly superior to ^{2}H coding. A second major advantage of succinate labeling is that the number of peptides labeled is much broader. All peptides in tryptic digests are labeled except those that are N-terminally blocked and contain no lysine.

20 CONCLUSION

It was observed during reversed-phase chromatography of peptides from a BSA tryptic digest differentially labeled with the $^{2}H_0$ and $^{2}H_8$-ICAT reagents that resolution of the isoforms exceeded 0.5 with 20% of the peptides in the digest. Three fourth of the peptides in this group contained two cysteine

25 residues and were doubly labeled. Only 23% of the peptides labeled with a single ICAT residue had a resolution greater than 0.4.

The resolution of peptides differentially labeled with ^{13}C and ^{12}C-succinate never exceeded ±0.01, even in the case of peptides from the BSA digest labeled with two moles of succinate. Because this value is within the

30 limits of the method used to determine resolution, it was concluded the ^{13}C and ^{12}C coded isoform of labeled peptides did not resolve. Abundance ratio in the case of $^{13}C/^{12}C$ coding could be determined from a single mass spectrum taken

at any point in the elution profile. This enabled abundance ratio analysis to be completed early in the elution of a peptide from chromatography columns.

It is concluded that isotopic fractionation can be minimized in isotope based, internal standard methods for quantifying relative concentrations of

5   peptides by using coding reagents containing ^{13}C instead of ^2H. As a result of eliminating isotope effects in reversed phase chromatography, quantification can be achieved much faster and with greater accuracy. This is expected to improve throughput and pave the way for real-time monitoring and intelligent data analysis.

10

Table 6. Comparison of resolution (R) caused by succinic anhydride-$^{13}C_4$ and ICAT-$^2H_8$.

peptide sequence	R (succinic anhydride-$^{13}C_4$)	number of ^{13}C	R (ICAT-$^2H_8$)	number of ^2H
CASIQK (SEQ ID NO:6)	-0.0081	8	0.32	8
GACLLPK (SEQ ID NO:7)	0.0012	8	0.41	8
LCVLHEK (SEQ ID NO:8)	0.0015	8	0.4	8
QNCDQFEK (SEQ ID NO:1)	-0.0043	8	0.74	8
SHCIAEVEK (SEQ ID NO:9)	0.0044	8	0.15	8
SLHTLFGDELCK (SEQ ID NO:10)	0.0091	8	0.27	8
YICDNQDTISSK (SEQ ID NO:11)	0.004	8	0.33	8
LFTFHADICTLPDTEK (SEQ ID NO:12)	-0.0045	8	0.28	8
GLVLIAFSQYLQQCPFDEHVK (SEQ ID NO:13)	0.0083	8	0.18	8
MPCTEDYLSLILNR (SEQ ID NO:14)	-0.0025	4	0.21	8

15

Example 3.

High Throughput Proteomics by Minimizing Isotopic Fractionation and Intelligent Data Acquisition (IDA)

5   This work further demonstrates that the isotopic fractionation can be minimized by using reagents containing ^{13}C or ^{18}O instead of ^{2}H, which in turn allows accurate quantification of relative protein expression/modification levels and high throughput identification of up or down-regulated proteins by real time ratio determination (RTRD) and intelligent data acquisition (IDA). An accurate

10   ratio between experimental and control samples can be determined in the early stage of their elution peak, and only those peptides that are significantly up or down-regulated need to be fragmented.

MATERIALS AND METHODS

15   *Materials.* Cytochrome c, HPLC grade acetonitrile (ACN), *N*-hydroxysuccinimide, ammonium bicarbonate, succinic anhydride, acetic anhydride, acetyl chloride, phosphorus pentoxide, urea, hydroxylamine hydrochloride were purchased from Sigma-Aldrich (St. Louis, MO). Trifluoroacetic acid (Sequanal Grade) was purchased from Pierce (Rockford,

20   IL). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Acetic-$^{2}H_6$-anhydride, succinic-$^{2}H_4$-anhydride, ^{18}O enriched water (95% enrichment), butanedioic-$^{13}C_2$-acid were purchased from Isotec (Miamisburg, OH). C18 column (4.6 x 250 mm) was purchased from Vydac (Hesperia, CA). Double diionized water (ddI H_2O) was produced by a Milli-

25   Q™ Gradient™ A10 System from Millipore (Bedford, MA).

      Proteolysis of cytochrome c. Cytochrome c (2 mg/mL) was mixed with sequencing grade modified trypsin at ratio of 50:1 (w/w). Proteolysis was carried out in 50 mM ammonium bicarbonate buffer for 12 hours at 37°C and stopped by freezing the mixture in liquid nitrogen for 10 minutes. Proteolysis

30   was performed in ^{18}O enriched water (95% enrichment) in the ^{18}O labeling experiments.

Synthesis of N-acetoxy-$^2H_3$-succinimide. A solution of 4.0 g of N-hydroxysuccinimide in 11.4 g of acetic-$^2H_6$-anhydride was stirred at room temperature. White crystals began to deposit in 10 minutes. After 15 hour the solution was filtered. The crystals were washed with hexane and then dried in

5   vacuum. N-acetoxy-succinimide (synthesized by Peiran Liu), propionate-N-hydroxysuccinimide ester and propionate-$^2H_5$-N-hydroxysuccinimide ester (synthesized by Li Xiong) were synthesized in a similar way.

Synthesis of succinic-$^{13}C_2$-anhydride. 2.0 g of butanedioic-$^{13}C_2$-acid and 5 mL of acetyl chloride were heated under reflux for 1.5 hours. After reaction

10 the clear solution was cooled to room temperature and kept at 0 °C overnight. The crystals formed upon cooling was further purified from dry diethyl ether and dried in *vacuo* over P_2O_5.

Derivatization of peptides. A fifty fold molar excess of derivatization reagents (i.e. succinic anhydride, succinic-$^{13}C_2$-anhydride, succinic-$^2H_4$-

15 anhydride, N-acetoxysuccinimide, N-acetoxy-$^2H_3$-succinimide, propionate-N-hydroxysuccinimide ester and propionate-$^2H_5$-N-hydroxysuccinimide ester) were added individually to experimental and control samples in ammonium bicarbonate buffer. Labeling reagents were added in small aliquots over the course of the first hour and the reaction was allowed to proceed for another two

20 hours. N-hydroxylamine was then added in excess and the pH was adjusted to 11-12. Incubation with hydroxylamine was allowed to proceed for 10 min to hydrolyze esters that might have been formed.

Reversed-phase elution of isotopically labeled peptides. Isotopically labeled peptide mixtures were separated by gradient elution from a Vydac C18

25 column on an Integral Micro-Analytical Workstation (Applied Biosystems, Framingham, MA). The C18 column was equilibrated using 100% mobile phase A (0.01% TFA in ddI H_2O) at a flow rate of 1.0 mL/min for 2 column volumes (CV). Isotopically labeled peptide mixtures (2 nmol) were injected and gradient elution of the analytes was achieved using 100% mobile phase A

30 to 60% mobile phase B (95% ACN/0.01% TFA in ddI H_2O) over 60 minutes, then 60% B to 100% B in 10 minutes at a flow rate of 1.00 mL/min. The gradient was then held at 100% mobile phase B for an additional 10 minutes. Throughout the analysis an on-line UV detector set at 214 nm was used to

monitor separation of the peptide mixtures. The peptides were simultaneously monitored by ESI-MS by directing 5% of the flow to the ion source.

    ESI-MS analysis. Mass spectral analyses were performed using a QSTAR workstation (Applied Biosystems, Framingham, MA) equipped with an

5   Ionspray source. All spectra were obtained in the positive TOF mode at a sampling rate of one spectrum every two seconds. During the LC-MS acquisition, masses were scanned from m/z 300 to 1500.

RESULTS AND DISCUSSION

10   The goals of this study are to (1) determine the degrees of isotopic fractionation of peptides derivatized by various reagents containing ^2H, ^{13}C or ^{18}O, (2) illustrate the problems when isotopic fractionation is significant, (3) demonstrate that isotopic fractionation can be minimized by using reagents containing ^{13}C or ^{18}O instead of ^2H, which then allows real time ratio

15   determination (RTRD) and intelligent data acquisition (IDA).

    Tryptic peptides from cytochrome c were isotopically labeled, mixed and subjected to LC-MS separation and quantification. Figure 11 shows mass spectra of a peptide (IFVQK; SEQ ID NO:2) labeled by succinic anhydride (m/z=834.4, control sample), succinic-$^{13}C_2$-anhydride (m/z=838.4, experimental

20   sample 1) and succinic-$^2H_4$-anhydride (m/z=842.4, experimental sample 2) at different time. Four ^{13}C and eight ^2H were incorporated in experimental sample 1 and experimental sample 2 respectively because both the amine group at the amino terminus and the lysine at the carboxyl terminus were derivatized by succinic anhydride. Figure 11 demonstrates that the ratio between experimental

25   sample 1 and control sample does not change with time while the ratio between experimental sample 2 and control sample varies continuously with time.

    A more quantitative analysis is seen in Figure 12. Figure 12 shows the extracted ion chromatograms of monoisotopic peaks at m/z of 834.4 (control sample), 838.4 (experimental sample 1) and 842.4 (experimental sample 2) and

30   the ratios between experimental samples and control sample. The continuously decreasing ratio between the ^2H labeled experimental sample 2 and the non-isotope labeled control sample indicates that the ^2H labeled peptide eluted significantly earlier than the non-isotope labeled control sample. The ^2H

labeling even changed the shape of the chromatographic peak which can be easily visualized and is also indicated by the increase of the ratio late in the chromatographic peak.

On the contrary to ^2H, no significant isotopic fractionation was observed

5   between the ^{13}C labeled peptide and non-isotope labeled peptide, as the ratio between them was constant across the chromatographic peak (Figure 12).

The same analysis was done to the individual isotope peaks (m/z=835.4, 836.4, 837.4, Figure 11), which showed no significant separation from the monoisotopic peak (m/z=834.4, Figure 11). One reason could be that the

10   isotope peaks at m/z of 835.4, 836.4 and 837.4 mainly consist of ^{13}C, etc. but not ^2H because the natural abundance of ^2H (0.015%) is much lower than that of ^{13}C (1.10%) etc.. Another reason could be that the heavy isotopes (^{13}C etc.) randomly distribute in the whole peptide instead of being enriched in one particular functional group. The insignificant fractionation between the isotope

15   peaks and the corresponding monoisotopic peak indicates that they can be added together to calculate the ratio between experimental and control samples to improve the precision of the ratio measurement.

As in Example 2, resolution (R) is defined as

$$R = \frac{\Delta}{W_{1/2}} \tag{11}$$

20   where Δ is the separation between the heavy isotope labeled experimental samples and the non-isotope labeled control sample and $W_{1/2}$ is the average full peak width at half maximum (FWHM) of the extracted ion chromatographic peaks of experimental and control samples.

Peptides labeled by propionate-$^2H_5$-N-hydroxysuccinimide ester showed

25   the greatest resolution (R) among all the five reagents tested (Figure 13A). The resolution is so large that most of the time the deuterated and non-deuterated peptides do not even show up in the same mass spectrum! When fractions are collected in the LC separation of a complicated peptide mixture, it is very probable that the deuterated and non-deuterated peptides go to two different

30   fractions so that the isotope ratio cannot be measured at all. Instead they may be mistakenly regarded as grossly up or down-regulated peptides since they appear as single isotope cluster instead of two isotope clusters with certain mass

difference. The popular ICAT reagent causes equal or larger R as eight or more deuterium is incorporated depending on the number of cystine the peptide has.

Resolution (R) caused by succinic-^{2}H$_4$-anhydride (Figure 13B) and N-acetoxy-^{2}H$_3$-succinimide (Figure 13C) are comparable to each other and much

5   smaller than that of propionate-^{2}H$_5$-N-hydroxysuccinimide ester while they are still significant. When fractions are collected to do MALDI-MS or nanospray-MS, they could be enriched in either the deuterated or non-deuterated peptides so that the isotope ratios are not accurate. In online LC-MS experiments ratio may be estimated in two ways. One way is to calculate the ratio based on

10   integration of the extracted ion chromatographic peaks of individual isotopically labeled peptides, which can only be done after the peptides completely elute from column so that real time ratio determination (RTRD) and intelligent data acquisition (IDA) are impossible. A better way is to keep tracking the intensities of elution peaks of deuterated and non-deuterated peptides. Once the

15   non-deuterated peptide reaches peak maximum, the ratio is estimated based on integrated peak area of the leading half of the non-deuterated and deuterated peptides. The peptides are fragmented in the trailing half of the elution peaks if the ratio is significantly different from one. It works when peak shapes of the isotopically labeled peptides are the same, which is not always true for the ^{2}H

20   labeling reagents (e.g. Figure 13).

^{18}O can be incorporated into peptides during trypsin digestion and be used in quantification. ^{18}O labeling could be a valuable alternative to amino acylation strategies. However, important issues in case of ^{18}O labeling are that (1) more than one ^{18}O can be incorporated into a peptide, (2) incorporation

25   varies with structure, and (3) peptides arising from the C-terminus of a protein are not labeled.

Incorporation of multiple molecules of ^{18}O into a peptide is a unique property of trypsin as opposed to other proteolytic enzymes. Apparently, trypsin covalently associates reversibly with peptides containing basic C-

30   terminal amino acids. Exchange of ^{18}O into these peptides even occurs in the absence of proteolysis when H$_2$^{18}O is added to a trypsin digest still containing trypsin. We found that this reaction is sometimes inhibited by urea. When urea is eliminated, in many cases two molecules of ^{18}O are incorporated in pure

71

$H_2{}^{18}O$. Again resolution varied between peptides with the greatest resolution occurring in smaller peptides.

In this example, resolution caused by ^{18}O (Figure 13D) was smaller than that of succinic-$^2H_4$-anhydride and N-acetoxy-$^2H_3$-succinimide. However, it

5   can still cause 15% or more errors when fractions are collected in the worst cases. Also the resolution is still not small enough to allow real time ratio determination (RTRD) from single mass spectrum snapshot because, for example, in Figure 13D ratio varied from 1.3 to 0.6 across the elution peak when the true value is 0.87. In this case real time ratio determination (RTRD)

10   works well using the integration of the leading halves of the elution peaks of isotopically labeled peptides because now R is sufficiently small and peak shapes are identical.

Succinic-$^{13}C_2$-anhydride showed the smallest resolution among all five reagents tested in this work. In Figure 13E, ratio varied less than 5% across the

15   elution peak. Now systematic errors are negligible (less than 1%) when fractions are collected. Real time ratio determination (RTRD) can be performed much earlier before reaching the peak maximums, which leaves more time for optimizing conditions and accumulate high quality MS/MS spectra.

Also minimizing R makes the accurate isotope ratios preserved in the

20   fragment ions when the mass windows are wide enough that two isotope clusters are simultaneously fragmented. This is particularly useful when multiple pairs of peptides overlap in mass spectra. It is difficult to do deconvolution to obtain the accurate isotope ratios of each pair of peptides. In this case, the only hope is that different pairs of peptides generate different

25   fragment ions and some of the fragment ions do not overlap with others and thus allow the calculation of accurate isotope ratios of original pair of peptides. The ratio calculated based on fragment ion clusters in MS/MS spectra reflect the ratio at the point that the fragmentation takes place, which could be very different from the true value when R is not small and ratio varies across the

30   elution peak.

Figure 14 shows that the larger the peptide mass the smaller the resolution (R). Also in all the cases of ^2H labeling, peptides end with lysine are derivatized twice, have twice as many ^2H as peptides end with arginine, and

have a correspondingly larger R when peptide masses are similar. Even resolution of peptides having similar masses and are derivatized by the same reagent with the same amount of ^{2}H or ^{18}O incorporation still varies, which means resolution also depend on peptide sequences/structures. However,

5   Resolution is independent of peptide retention time as resolution is dependent of peptide mass and retention time is independent of peptide mass. In the case of ^{13}C labeling no clear pattern was found probably because R is already within measurement error.

Figure 15 again shows the decrease of resolution in the order of

10   propionate-^{2}H$_5$-N-hydroxysuccinimide ester, succinic-^{2}H$_4$-anhydride, N-acetoxy-^{2}H$_3$-succinimide, H$_2$^{18}O, succinic-^{13}C$_2$-anhydride in the cases of three peptides from cytochrome c. Resolution (R) should also depend on column and gradient program.

An interesting fact is that the natural abundance of ^{2}H (0.015%) is much

15   lower than that of ^{13}C (1.10%), ^{18}O (0.2%) or ^{15}N (0.37%). However, reagents containing ^{2}H are much cheaper than that containing ^{13}C, ^{15}N, ^{18}O etc. The reason may be that the difference between ^{2}H and ^{1}H is much larger than that between ^{13}C and ^{12}C, ^{15}N and ^{14}N, ^{18}O and ^{16}O, etc. and thus ^{2}H is the easiest heavy isotope to purify. At the mean time the large difference between ^{2}H and

20   ^{1}H makes deuterated compounds not the best internal standards.

Another fact is that the cost of ^{13}C containing reagents is not very high considering only several milligrams of reagents is used in each experiment, while gains in accuracy and throughput are very significant.

25   Example 4.
         Chemically Multiplexed Global Internal Standard Technology (mGIST)

Elimination of isotope effect makes multiplexing possible.

         The global internal standard technique (GIST) (PCT WO 01/86306,

30   published November 15, 2001; U.S. patent application number 09/849,924) allows 1) quantification in the cases of both protein expression and post-translational modifications as they relate to cellular regulation and 2) quantification may be achieve with all types of targeting schemes. Through

various kinds of group specific chromatographic selectors, it is possible to target natural features of peptides and proteins for selection and subsequent analysis. For example, in the case of protein expression, either histidine- or cysteine-containing peptides or those that contain both of these amino acids can be

5 selected. (Surprisingly, almost as many proteins produce a tryptic peptide carrying both histidine and cysteine as either amino acid alone. The great advantage of double selection is that there are so many fewer peptides with both amino acids.) As another example, post-translational modifications such as glycosylation and phosphorylation can also be used for selection.

10   There remains the problem of increasing throughput. A number of recent presentations have addressed the throughput issue with what we call the "Gatling gun" or physical multiplexing approach. Although there are many clever ways to achieve parallel processing with LC/MS instrumentation, all the reports have a common thread. Separations are carried out with multiple

15 columns or gels simultaneously and fractions from each column or gel are either 1) mechanically collected or transferred directly onto a MALDI plate for later analysis in a single mass spectrometer, 2) electrosprayed into a single mass spectrometer through valve switching, or 3) electrosprayed into one of an array of mass spectrometers. The requisite replication and synchronization of

20 multiple components can be both costly and complex. Equally important is the fact that the total number of column fractions (F) generated by a mechanically multiplexed systems is equal to the number being generated from a single column times the number of columns (N), i.e. $F = fN$. Because a mass spectrum must be generated for each of these fractions, the total number of mass spectra

25 (S) taken will be directly proportional to F. This means that mechanical multiplexing neither reduces the number of spectra that must be taken nor the amount of sample processing that must be done. S is the same in all cases. Only the number of samples delivered per unit time changes.

      An alternative strategy to increase throughput in proteomics using

30 "chemical multiplexing" (mGIST) is possible. mGIST means synthesizing multiple reagents which are chemically identical but have different numbers of isotopes. We can then use the reagents to "code" multiple samples simultaneously. We already know that 1) the global derivatization strategies

74

described above make it possible to code every peptide in a sample without regard to sample complexity, 2) identical peptides from two samples can be traced through complex analyses involving many steps, 3) when a coded peptide has been identified or sequenced, the sequence of isoforms is also known, and

5   4) relative concentration of an analyte in two samples can be determined when both are analyzed simultaneously. Exploiting these facts could substantially increase throughput. It is expected that this strategy can be expanded to allow 3-5 samples to be analyzed simultaneously.

        The concept underlying mGIST is to fill a portion of the unused

10   separation space in mass spectra with isotopically coded forms of peptides from multiple samples, at least until multiple components fill the same separation space, i.e. isobaric peptides are being generated. In mGIST the abundance ratio should be well defined. For example, it is preferred that the abundance ratio be normalized against the most abundant analyte or against a control sample.

15   This chemical multiplexing approach has many advantages. One is that the number of column fractions being generated is independent of the number of samples being analyzed. Another is that more of the analytical work is being loaded onto the faster, higher resolution mass spectrometer. Still another is that increases in the amount of data processing required are far less than the increase

20   in the number of samples. The fact that throughput would be increased with existing systems is also attractive. And finally, quantitative comparisons between multiple samples could easily be achieved.

        Importantly, limiting the resolution of a differentially derivatized peptides as in the preceding Examples enables sample multiplexing during

25   analysis. For example, differential derivatization of samples with a series of isoforms of an isotope coding agent that differ in mass by three daltons (e.g., isoforms having masses of 3, 6 and 9 daltons in excess of the mass of the isoform containing no heavy isotopes) will allow at least four samples to be analyzed concurrently. (For convenience, these heavy isotope isoforms can be

30   generically referred to as ^{3}CODR, ^{6}CODR and ^{9}CODR, respectively, with ^{0}CODR referring to isoform that contains no heavy isotopes, and where CODR stands for the _cod_ing _r_eagent.) This means that individual components in 3-5 samples can be chemically coded and quantified simultaneously using a single

analytical system. Targeted, quantitative proteomics thereby become amenable
to chemically multiplexed global internal standard technology (mGIST).

Oxygen-18 mGIST coding reagents. Oxygen-18 has certain features that
diminish its attractiveness for mGIST reagents. One is that mass would only

5 increases two amu per oxygen atom added to the coding reagent. A second
problem is the abundance issue in organic compounds. Although it is possible
to synthesize a set of four mGIST reagents varying by a single ^{18}O, a correction
would have to be made for their overlap with M+2 isotope peaks from natural
isotopes containing ^{13}C and ^{2}H. Corrections work well when the $^{2}CODR$

10 derivatized peak is equal to or larger than the $^{0}CODR$ peak, but become more
problematic as the $^{2}CODR$ derivatized peak becomes smaller relative to the
$^{0}CODR$ peak. Distinguishing the $^{2}CODR$ peptide peak from the M+2 peak of
the $^{0}CODR$ peptide is difficult when the ($^{0}CODR$ peptide)/($^{2}CODR$ peptide)
ratio is high.

15 It is desirable that coding reagents vary by three or more atomic mass
units (amu) to circumvent the problem of overlap with natural isotope peaks.
This suggests that coding reagents based on ^{18}O should vary by two ^{18}O atoms.
At least six ^{18}O atoms would be required to produce a set of four coding
reagents varying by four amu. Carbohydrates and polyoxyethylenes meet this

20 criterion, although the requisite isotopically labeled derivatizing reagents are not
at present commercially available.

Carbon-13 coding reagents. It is desirable that the mass difference
between coding reagents be at least three amu, production of a 0, +3, +6, and +9
amu set would require a CODR structure with at least nine carbon atoms that

25 could be manipulated.

m-GIST in the study of post-translational modifications. Most of the
work in proteomics related to post-translational modification (PTM) is along
two lines. One is to use unique mass spectral characteristics to identify
particular types of PTM, such as in the case of phosphorylation. This is a

30 lengthy process. Another is to use affinity chromatography systems to select
particular types of PTM as we have done with glycosylation and
phosphorylation (PCT WO 01/86306, published November 15, 2001). The
problem is that affinity selectors may not be available for all types of PTM.

76

mGIST enables another approach. There are a number of enzymes that cleave post-translational modifications from peptides or proteins. The specificity of these enzymes can be used in an analytical mode for PTM identification. The control sample is split into two fractions and one is treated

5   with the analytical enzyme. Following tryptic digestion of the two fractions, the untreated control would be derivatized with a ^{0}CODR and the enzyme treated sample with the ^{6}CODR species. The experimental sample would be treated with trypsin and derivatized with the ^{3}CODR. All three samples will then be mixed and fractionated.

10   This strategy leads to the experimental sample being boxed between an untreated and enzyme treated control. This means all peptide peaks will appear as triplet clusters in mass spectra with one exception. Those in which the enzyme cleaved a PTM from the protein will cause peptides to appear as doublet clusters. This unique pattern will be readily apparent and identify the

15   PTM. The first and second clusters may still be used for quantification and location of the PTM by MS/MS sequencing.

This concept can be tested with alkaline phosphatase to detect phosphorylation, low molecular weight tyrosine phosphatase to search for tyrosine phosphorylation, serine/threonine phosphatase, sulfatases,

20   glycosidases, and a variety of other enzymes to detect hydrolysis of their respective substrates.

The complete disclosures of all patents, patent applications including provisional patent applications, and publications, and electronically available

25   material (e.g., GenBank amino acid and nucleotide sequence submissions) cited herein are incorporated by reference. The foregoing detailed description and examples have been provided for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described; many variations will be apparent to one

30   skilled in the art and are intended to be included within the invention defined by the claims.

SELECTED BIBLIOGRAPHY

S. P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M. Gelb, and R. Aebersold. Quantitative analysis of complex proteins mixtures using isotope-coded affinity tags. Nat. Biotechnol. (1999) 17, 994-999.

J. Ji, A. Chakraborty, M.Geng, X. Zhang, A. Amini, M. Bina, and

5   F.Regnier. Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J. Chromatogr. B,* (2000) 745, 197-210.

X. Yao, A. Freas, J. Ramierez, P. Demirev, and C. Fenselau. Proteolytic [18]O labeling for comparative proteomics: Model studies with two serotypes of adenovirus. Anal. Chem. (2001) 73, 2836-2842.

10   T. P. Conrads, K. Alving, T.D. Veenstra, M. Belov, G. Anderson, D. Anderson, M. Lipton, L. Pasa-Tolic, J. Udseth, W, Chrisler, B. Thrall, and R. Smith. Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N metabolic labeling. Anal. Chem. (2001) 73, 2132-2139.

15   D.Goodlett, A.Keller, J.Watts, R.Newitt, E.Yi, S.Purvine, J.Eng, P, Von Haller, and R.Aebersold. Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. Rapid Commun. Mass Spectrom. (2001) 15, 1214-1221.

R. Zhang, C. Sioma, S. Wang, and F. Regnier. Fractionation of

20   isotopically labeled peptides in quantitative proteomics. Analytical Chemistry, (2001) 73, 5142-5149.

A. Amini, S. Dormady, L. Riggs, Larry; and F.Regnier. The impact of buffers and surfactants from micellar electrokinetic chromatography on matrix-assisted laser desorption ionization (MALDI) mass spectrometry of peptides. J.

25   of Chromatogr. A , (2000) 894, 345-355.

F. Regnier, A. Amini, A. Chakraborty, M. Geng, J. Ji, L. Riggs, C. Sioma, S. Wang, and X. Zhang, Xiang. Multidimensional chromatography and the signature peptide approach to proteomics. LC/GC, (2001) 19, 200-213.

S. Wang and F. Regnier. Covalent chromatography for the selection of

30   cysteine containing peptides in proteomics. J. Chromatogr. (2001) 924(1-2), 345-357.

M. Larsen, P. Larsen, S. Fey, and P. Roepstroff. Characterization of differently processed forms of enolase 2 from *Saccharomyces cerevisiae* by

two-dimensional gel electrophoresis and mass spectrometry. Electrophoresis. (2001) 22, 566-575.

T. Kosaka, T. Takazawa, and T. Nakamura. Identification and C-terminal characterization of proteins from two-dimensional polyacrylamide gels
5   by a combination of isotope labeling and nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry Anal. Chem. (2000) 72, 1179-1185.

P.K. Jensen, L. Pasa-Tolic, G. Anderson, J. Homer, M. Lipton, J. Bruce, and R. Smith. Probing proteomics using capillary isoelectric focusing electrospray ionization Fourier transform ion cyclotron resonance mass
10   spectrometry. Anal. Chem. (1999) 71, 2076-2084.

O.A. Kirgorodskaya, Y.P. Koz'min, M.I. Titov, N.V. Savel'eva, R. Komer, C. Sonksen, A.I. Miroshnikov, P. Roepstorff. Quantitative determination of peptides and proteins by MALDI MS. Fuss. J. Biorg. Chem. (2000) 26, 593-602.
15   B. Kuster and M. Mann. ^{18}O labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. Anal. Chem. (1999) 71, 1431-1440.

J.D. Watson. The human genome project; past, present, and future. Science. (1990) 248, 44-49.
20   M. Washburn, D. Wolthers, and J. Yates. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. (2001) 19, 242-247.

M. Mann, R. Hendrickson, and A. Pandey. Analysis of proteins and proteomes by mass spectrometry. Ann. Rev. Biochem. (2001) 70, 437-473.
25   N. Anderson, A. Matheson, and L. Anderson. Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. Proteomics (2001) 1, 3-12.

B. Walsh, M. Molloy, and K. Williams. The Australian proteome analysis facility (APAF). Assembling large-scale proteomics through
30   integration and automation. Electrophoresis. (1998) 19, 1883-1890.

G. Cothals and P. Nelson. Large-scale proteomics and its future impact on medicine. Pharmagocgenomics J. (2001) 1, 15-19.

M. Van der Werf, F. Schuren, S. Bijlsma, A.C. Tas, and B. Van Ommen. Nutrigenomics: application of genomics technologies in nutritional sciences and food technology. J. Food Sci, (2001) 66, 772-780.

G. Thomas. Metabolomics breaks the silence. Trends Biotech. (2001) 19, 126-127.

O. Fiehn. Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks. Comp. Funct. Genomic. (2001) 2, 156-158.

L. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. Walsh, J. Berden, K. Brindle, D. Kell, J. Rowland, H. Westerhoff, K. Van Dan, and S. Oliver. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat. Biotechnol. (2001) 19, 45-50.

M. Foersma, I. Solyanikova, W. Van Berkel, L. Golovleva, and I. Rietjens. ^{19}F NMR metabolomics for the elucidation of microbial degradation pathways of fluorophenols. J. Ind. Microbiol. Biotechnol. (2001) 26, 22-34.

K.D.W.Roth, Z.-H.Huang, N. Sadagopan, and J.Throck Watson. Charge Derivatization of peptides for analysis by mass spectrometry. Mass Spectrometry. Rev. (1998) 17, 255-274.

L. Andersson. The use of immobilized Fe^{+3} and other hard metal ions in chromatography of peptides and proteins. Int. J. Bio-Chromatogr. (1996) 2, 25-36.

H. Kaufmann, J. Gailey, and M. Fussenegger. Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. Proteomics (2001) 1, 194-199.

M. Dwek, J. Ross, and A. Leathem. Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. Proteomics (2001) 1, 756-762.

J. Ji, A. Chakraborty, M. Geng, X. Zhang, A. Amini, M. Bina, and F. Regnier, Fred. Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J. Chromatogr. B,* (2000) 745, 197-210.

M. Geng, X. Zhang, M. Bina, and F. Regnier. Proteomics of glycoproteins based on affinity selection of glycopeptides from tryptic digests. J. Chromatogr. B, (2001) 752, 293-306.

L. Riggs, C. Sioma, and F. Regnier. An Automated Signature Peptide Approach for Proteomics. J. Chromatogr. A, (2001) 924(1-2), 359-368.

H. Wang, C. Desilets, and F. E. Regnier. A Study of Retention Mechanism in Semipermeable Surface Chromatography. Anal. Chem. (1992)
5   64, 2821-2827.

C. Dass, P. Mihalakshmi, and D. Gandberry. Manipulation of the ion-paining reagents for reversed-phase high-performance liquid chromatographic separation of phosphorylated opoid peptides from their non-phosphorylated analogues. J. Chromatogr. A (1994) 67, 249-257.

10   M. Schnoelzer, P. Jedrzejeski, W.D. Lehman. Protease-catalyzed incorporation of ^{18}O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. Electrophoresis. (1996) 17, 945-953.

B. Karger. High throughput mass spectrometry. Ab. Paper. Am. Chem.
15   Soc. (2001) 221st IEC-386.

R.D.Rocklin, R.Ramsey, J.M.Ramsey. A microfabricated fluidic device for performing two-dimensional liquid phase separation. Anal. Chem. (2001) 72, 5244-5249.

E.R. Badman and R.G.Cooks. Cylindrical ion trap array with mass
20   selection by variation in trap dimensions. Anal. Chem. (2000) 72, 5079-5086.

S.D. Patterson, R. Aebersold, and D. Goodlett. Mass spectrometry-based methods for protein identification and phosphorylation site analysis. Proteomics (2001) 1, 87-130.

N. Ishizuka, H. Minakuchi, K. Nakanishi, N, Soga, H. Nagayama, K,
25   Hosoya, and N. Tanaka. Performance of monolithic silica column in a capillary under pressure-driven and elctrodriven conditions. Anal. Chem. (2000) 72, 1275-1280.

Parker, K. C.; Griffin, T.; Gygi, S.; Aebersold, R. 48th ASMS Conference on Mass Spectrometry and Allied Topics, 2000.

30   S. Wang, X. Zhang, and F. Regnier. A quantitative proteomics strategy involving the selection of peptides containing both cysteine and histidine from tryptic digests of cell lysates. J. Chromatogr. A (2002) 949(1-2), 153-162.

WHAT IS CLAIMED IS:

1. An isotope coding agent comprising:

a reactive functional group that reacts with an amine, a thiol, a hydroxyl

5   or a carboxyl; and

an isotopic linker comprising at least three heavy non-deuterium

isotopes.

2. The isotope coding agent of claim 1 further comprising an affinity functional

10   group for affinity selection.

3. The isotope coding agent of claim 1 wherein each heavy non-deuterium

isotope is independently selected from the group consisting of ^{13}C, ^{18}O and ^{15}N.

15   4. The isotope coding agent of claim 1 wherein the heavy non-deuterium

isotopes are identical.

5. The isotope coding agent of claim 1 comprising at least three ^{13}C isotopes.

20   6. The isotope coding agent of claim 1 comprising at least three ^{18}O isotopes.

7. The isotope coding agent of claim 1 wherein at least two of the heavy non-

deuterium isotopes differ.

25   8. The isotope coding agent of claim 1 wherein the reactive functional group

iodoacetic acid or iodoacetamide and wherein the affinity functional group

biotin.

9. An isotope coding agent comprising:

30   a reactive functional group selected from the group consisting of

iodoacetic acid and iodoacetamide;

an isotopic linker having the structural formula $C_{10}H_{17}N_3O_3$ and

comprising about nine ^{13}C isotopes; and

a functional group for affinity selection comprising biotin.

10. An isotope coding agent comprising a compound selected from the group consisting of succinic anhydride, N-acetoxysuccinimide and propionate-N-
5   hydroxysuccinimide, wherein the compound comprises at least three heavy non-deuterium isotopes.

11. An isotope coding agent comprising at least three ^{13}C isotopes, wherein the isotope coding agent reacts with an amine, a carboxyl, a hydroxyl or a thiol.
10

12. An isotope coding agent comprising at least three ^{18}O isotopes, wherein the isotope coding agent reacts with an amine, a carboxyl, a hydroxyl or a thiol.

13. An isotope coding agent comprising at least three ^{15}N isotopes, wherein the
15   isotope coding agent reacts with an amine, a carboxyl, a hydroxyl or a thiol.

14. A method for making an isotope coding agent comprising incorporating at least three heavy isotopes independently selected from the group consisting of ^{13}C, ^{18}O and ^{15}N into a compound to yield the ^{13}C-, ^{18}O- and/or ^{15}N-containing
20   isotope coding agent, wherein a deuterated isoform of the compound is in use as a isotope coding agent in mass spectrometry.

15. The method of claim 14 wherein the isotope effect observed for the ^{13}C-, ^{18}O- and/or ^{15}N-containing isotope coding agent is smaller than the isotope
25   effect observed for the deuterated isoform.

16. A method for making an isotope coding agent comprising:
        identifying a deuterated isotope coding agent; and
        synthesizing an isoform of the denterated isotope coding agent, wherein
30   the isoform comprises at least three heavy isotopes independently selected from
the group consisting of ^{13}C, ^{18}O and ^{15}N.

17. The method of claim 16 wherein the isotope effect observed for the ^{13}C-, ^{18}O- or ^{15}N-containing isotope coding agent is smaller than the isotope effect observed for the deuterated isotope coding agent.

5   18. A peptide covalently linked to an isotope coding agent, wherein the isotope coding agent comprises at least three ^{13}C isotopes.

19. The peptide of claim 18 wherein the isotope coding agent is selected from the group consisting of an ICAT reagent, succinic anhydride, N-

10   acetoxysuccinimide and propionate-N-hydroxysuccinimide.

20. A peptide covalently linked to an isotope coding agent, wherein the isotope coding agent comprises at least three ^{18}O isotopes.

15   21. The peptide of claim 20 wherein the isotope coding agent is selected from the group consisting of an ICAT reagent, succinic anhydride, N-acetoxysuccinimide and propionate-N-hydroxysuccinimide.

22. A peptide covalently linked to at least two different isotope coding agents,

20   wherein each isotope coding agent comprises at least one heavy isotope independently selected from the group consisting of ^{13}C, ^{18}O and ^{15}N.

23. A method for isotopically coding an analyte comprising covalently linking the analyte to an isotope coding agent, wherein the isotope coding agent

25   comprises at least three heavy isotopes independently selected from the group consisting of ^{13}C, ^{18}O and ^{15}N.

24. A device for detecting a difference in the concentration of an analyte present in a first sample and in a second sample, the device comprising:

30   a sample fractionator having an outlet;

                 a mass spectrometer coupled to the outlet of the sample fractionator; and

software for determining the abundance ratio for the analyte using the mass spectrum of a combined sample immediately following elution of isotopically labeled analytes from the sample fractionator.

5   25. The device of claim 24 wherein the software determines a normalized abundance ratio characterizing analytes whose concentration is the same in the first and second samples and an abundance ratio of the first and second isotopically labeled analytes, and wherein a difference in the abundance ratio of the first and second isotopically labeled analytes and the normalized abundance
10   ratio is indicative of a difference in concentration of the analyte in the first and second samples.

26. The device of claim 24 wherein the abundance ratios are determined in real time such that an eluted fraction comprising analytes that have changed in
15   concentration can be immediately subjected to further analysis using MS-MS.

27. The device of claim 24 wherein the software is designed for proteomics applications.

20   28. The device of claim 24 wherein the software is designed for blood sample analysis.

29. The device of claim 24 wherein the analytes further analyzed on-line, in real time, using MS-MS.
25
30. The device of claim 24 wherein the decision of whether to perform MS-MS on a peak is made by the software without input from a human operator.

31. The device of claim 24 wherein the sample fractionator is a
30   chromatography column.

32. The device of claim 24 wherein the sample fractionator is a reverse phase column.

33. A method for using the device of claim 24 comprising:

    fractionating a combined sample using the fractionator, wherein each constituent sample comprises an isoform of each of a plurality of analytes, at

5   least one isoform of each analyte comprising an isotope coding reagent comprising at least one non-deuterium heavy atom and having a mass of at least 3 amu greater than the mass of an analyte isoform comprising no heavy isotopes, said fractionation yielding a plurality of elution peaks;

    performing MS on each elution peak;

10   determining the relative abundance ratio in real time of different isoforms of the analyte to detect isoforms exhibiting a change in concentration; and

    performing MS-MS on the elution peak comprising the isoforms exhibiting a change in concentration to further analyze said isoforms.

15

34. The method of claim 33 further comprising identifying the analyte exhibiting the change in concentration.

35. The method of claim 33 wherein the analytes are peptides produced by

20   proteolysis of samples containing proteins, and wherein MS-MS analysis of the elution peak comprising isoforms of peptides exhibiting a change in concentration is used to identify the proteins from which the peptides were derived.

25   36. The method of claim 33 wherein the isotope coding reagent contains no deuterium atoms.

37. The method of claim 33 wherein the isotope coding reagent comprises at least one heavy isotope selected from the group consisting ^{13}C, ^{18}O and ^{15}N.

30

38. The method of claim 33 wherein the different isoforms of an analyte do not resolve during fractionation.

39. A method for detecting a difference in the concentration of an analyte present in a first sample and in a second sample, each sample comprising a plurality of analytes, the method comprising:

covalently attaching a first isoform of a labeling agent to the analyte in
5   the first sample to yield at least one first isotopically labeled analyte;

covalently attaching a second isoform of the labeling agent to the analyte in the second sample to yield at least one second isotopically labeled analyte, wherein the first isoform comprises at least one non-deuterium heavy isotope, and wherein the first and second isoforms differ in mass by at least 3
10 amu;

mixing at least portions of the first and second samples to yield a combined sample; and

subjecting the combined sample to mass spectrometric analysis to determine a normalized abundance ratio characterizing analytes whose
15 concentration is the same in the first and second samples and an abundance ratio of the first and second isotopically labeled analytes, wherein a difference in the abundance ratio of the first and second isotopically labeled analytes and the normalized abundance ratio is indicative of a difference in concentration of the analyte in the first and second samples.

20

40. The method of claim 39 wherein the analyte is a protein.

41. The method of claim 39 further comprising fractionating the combined sample using a sample fractionator to yield at least one fraction comprising the
25 first and second isotopically labeled analytes prior to determining the abundance ratios.

42. The method of claim 41 wherein the isotope effect observed during fractionation of a mixture of the first and second isotopically labeled analytes is
30 smaller than the isotope effect observed during fractionation of a mixture comprising the second isotopically labeled analyte and an analyte labeled with a deuterated isoform of the labeling agent.

43. The method of claim 41 wherein fractionating the combined sample comprises subjecting the proteins to single or multi-dimensional reversed phase chromatography, ion exchange chromatography, hydrophobic interaction chromatography, size exclusion chromatography, capillary gel electrophoresis,

5   capillary zone electrophoresis, and capillary electrochromatography, capillary isoelectric focusing, immobilized metal affinity chromatography, affinity electrophoresis or any combination thereof.

44. The method of claim 41 wherein mass spectrometric analysis is performed

10   on the fraction immediately following elution of the fraction from the sample fractionator, and wherein abundance ratios are determined in real time such that an eluted fraction comprising analytes that have changed in concentration can be immediately subjected to further analysis using MS-MS.

15   45. The method of claim 38 wherein mass spectrometric analysis is performed using matrix assisted laser desorption ionization (MALDI), electrospray ionization (ESI), fast atom bombardment (FAB), electron impact ionization, atmospheric pressure chemical ionization (APCI), time-of-flight (TOF), quadrapole, ion trap, magnetic sector, ion cyclotron resonance mass, or

20   combinations thereof.

46. A method for detecting a difference in the concentration of a protein present in a first sample and in a second sample, each sample comprising a plurality of proteins, the method comprising:

25   covalently attaching a first isoform of a labeling agent to the protein in the first sample to yield at least one first isotopically labeled protein;

covalently attaching a second isoform of the labeling agent to the protein in the second sample to yield at least one second isotopically labeled protein, wherein the first isoform comprises at least one non-deuterium heavy isotope,

30   and wherein the first and second isoforms differ in mass by at least 3 amu;

cleaving proteins in the first and second samples to yield first and second isotopically labeled peptides in the first and second samples, respectively;

mixing at least portions of the first and second samples to yield a combined sample, wherein mixing is performed before or after fragmentation; and

subjecting the combined sample to mass spectrometric analysis to

5　determine a normalized abundance ratio characterizing peptides derived from proteins whose concentration is the same in the first and second samples and an abundance ratio of the first and second isotopically labeled peptides, wherein a difference in the abundance ratio of the first and second isotopically labeled peptides and the normalized abundance ratio is indicative of a difference in

10　concentration in the first and second samples of a protein from which the peptide is derived.

47. A method for detecting a difference in the concentration of a protein present in a first sample and in a second sample, each sample comprising a

15　plurality of proteins, the method comprising:

cleaving proteins in the first and second samples to yield at least one peptide in each sample;

covalently attaching a first isoform of a labeling agent to a peptide in the first sample to yield at least one first isotopically labeled peptide;

20　covalently attaching a second isoform of the labeling agent to a peptide in the second sample to yield at least one second isotopically labeled peptide, wherein the first isoform comprises at least one non-deuterium heavy isotope, and wherein the first and second isoforms differ in mass by at least 3 amu;

mixing at least portions of the first and second samples to yield a

25　combined sample; and

subjecting the combined sample to mass spectrometric analysis to determine a normalized abundance ratio characterizing peptides derived from proteins whose concentration is the same in the first and second samples and an abundance ratio of the first and second isotopically labeled peptides, wherein a

30　difference in the abundance ratio of the first and second isotopically labeled peptides and the normalized abundance ratio is indicative of a difference in concentration in the first and second samples of a protein from which the peptide is derived.

89

48. The method of claim 47 wherein the first and second isoforms of the labeling agent are attached to at least one amino group on peptides in the first and second samples.

5

49. The method of claim 47 wherein the first and second isotopically labeled peptides comprise at least one affinity ligand, the method further comprising, prior to determining the abundance ratios, contacting the peptides with a capture moiety to select peptides comprising the at least one affinity ligand.

10

50. The method of claim 47 further comprising identifying the protein from which the detected peptide was derived.

51. The method of claim 47 further comprising fractionating the combined

15   sample using a sample fractionator to yield at least one fraction comprising the first and second isotopically labeled analytes prior to determining the abundance ratios.

52. The method of claim 51 wherein the isotope effect observed during

20   fractionation of a mixture of the first and second isotopically labeled peptides is smaller than the isotope effect observed during fractionation of a mixture comprising the second isotopically labeled peptide and a peptide labeled with a deuterated isoform of the labeling agent.

25   53. The method of claim 51 wherein fractionating the combined sample comprises subjecting the proteins to single or multi-dimensional reversed phase chromatography, ion exchange chromatography, hydrophobic interaction chromatography, size exclusion chromatography, capillary gel electrophoresis, capillary zone electrophoresis, and capillary electrochromatography, capillary

30   isoelectric focusing, immobilized metal affinity chromatography, affinity electrophoresis or any combination thereof.

90

54. The method of claim 51 wherein mass spectrometric analysis is performed on the fraction immediately following elution of the fraction from the sample fractionator, and wherein abundance ratios are determined in real time such that an eluted fraction comprising analytes that have changed in concentration can

5   be immediately subjected to further analysis using MS-MS.

55. The method of claim 47 wherein the sample comprises at least about 100 proteins.

10   56. The method of claim 47 wherein the first sample is obtained from an environment before application of a stimulus and the second sample is obtained from an environment after application of the stimulus.

57. The method of claim 47 wherein the first sample is obtained from an

15   environment in the presence of, or after application of, a stimulus and the second sample is obtained from an environment in the absence of the stimulus.

58. The method of claim 47 wherein the first and second samples are obtained from different organisms, cells, organs, tissues or bodily fluids, the method

20   further comprising determining differences in concentration of at least one protein in the organisms, cells, organs, tissues or bodily fluids from which the samples were obtained.

59. A method for detecting a difference in the concentration of an analyte of

25   interest present in a multiplicity of samples, each sample comprising a plurality of analytes, the method comprising:

   for each sample, covalently attaching an isoform of a labeling agent to the analyte of interest to yield an isotopically labeled analyte of interest, wherein different isoforms of the labeling agent are used for each sample, the

30   isoforms differ from one another in mass by at least 3 amu, and all but at most one isoform comprise at least one non-deuterium heavy isotope;

   mixing at least portions of the samples to yield a combined sample; and

91

subjecting the combined sample to mass spectrometric analysis to
determine a normalized abundance ratio characterizing analytes whose
concentration is the same in the samples and an abundance ratio of at least one
pair of the isotopically labeled analytes of interest, wherein a difference in the
5   abundance ratio of the isotopically labeled analytes of interest and the
normalized abundance ratio is indicative of a difference in concentration of the
analyte of interest in the samples.

60. The method of claim 59 wherein the analyte of interest is present in about 3
10   to about 5 samples.

61. A method for detecting a difference in the concentration of a protein of
interest present in a multiplicity of samples, each sample comprising a plurality
of proteins, the method comprising:
15   cleaving the proteins in the samples to yield at least one peptide of
interest derived from the protein of interest;
            for each sample, covalently attaching an isoform of a labeling agent to
the peptide of interest to yield an isotopically labeled peptide of interest,
wherein different isoforms of the labeling agent are used for each sample, the
20   isoforms differ from one another in mass by at least 3 amu, and all but at most
one isoform comprise at least one non-deuterium heavy isotope;
            mixing at least portions of the samples to yield a combined sample; and
            subjecting the combined sample to mass spectrometric analysis to
determine a normalized abundance ratio characterizing peptides derived from
25   proteins whose concentration is the same in the first and second samples and an
abundance ratio of at least one pair of the isotopically labeled peptides of
interest, wherein a difference in the abundance ratio of the isotopically labeled
peptides of interest and the normalized abundance ratio is indicative of a
difference in concentration in the samples of a protein from which the peptide is
30   derived.

62. The method of claim 61 wherein the protein of interest is present in about 3
to about 5 samples.

63. A method for identifying isoforms of an analyte, the method comprising:

    fractionating a combined sample, wherein each constituent sample comprises an isoform of each of a plurality of analytes, at least one isoform of

5   each analyte comprising an isotope coding reagent comprising at least one non-deuterium heavy atom and having a mass of at least 3 amu greater than the mass of an analyte isoform comprising no heavy isotopes, and wherein the isoforms of the analyte do not resolve during fractionation, said fractionation yielding a plurality of elution peaks;

10   performing MS on each elution peak to detect isoforms of the analyte.

A

B

C
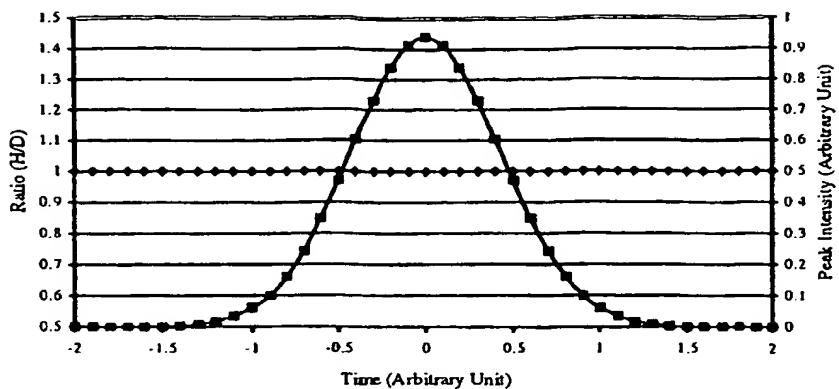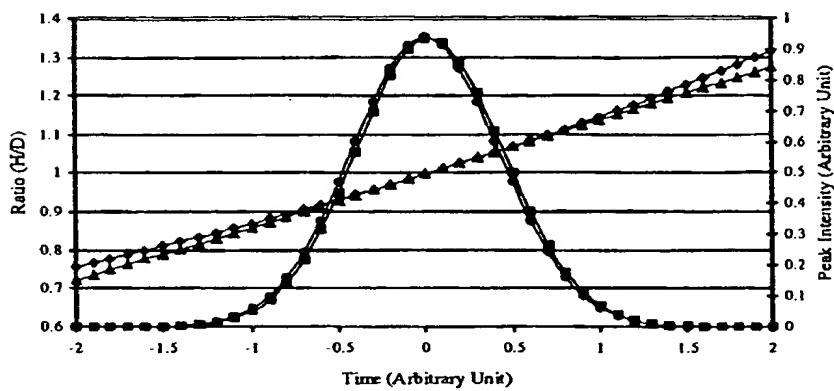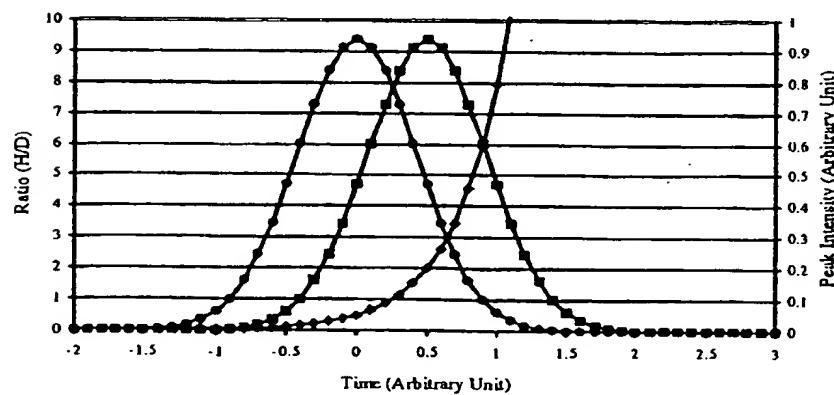
Fig. 1
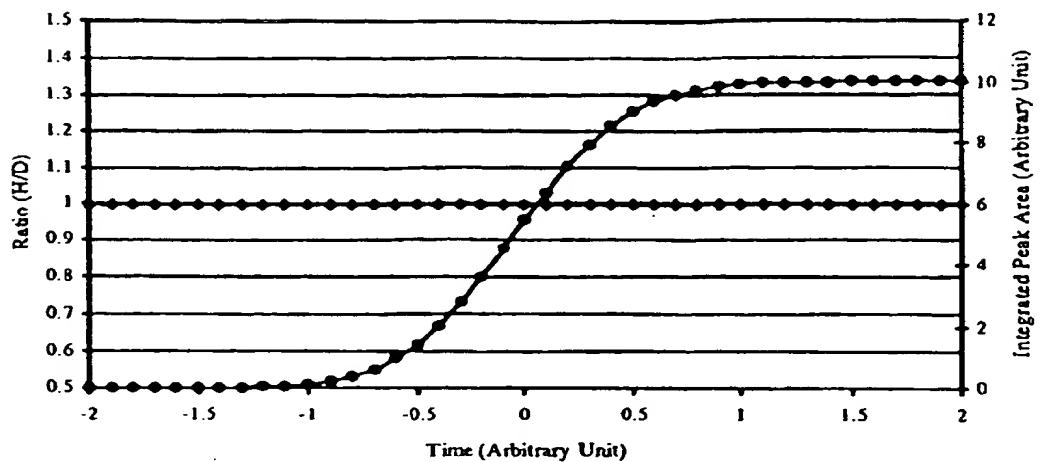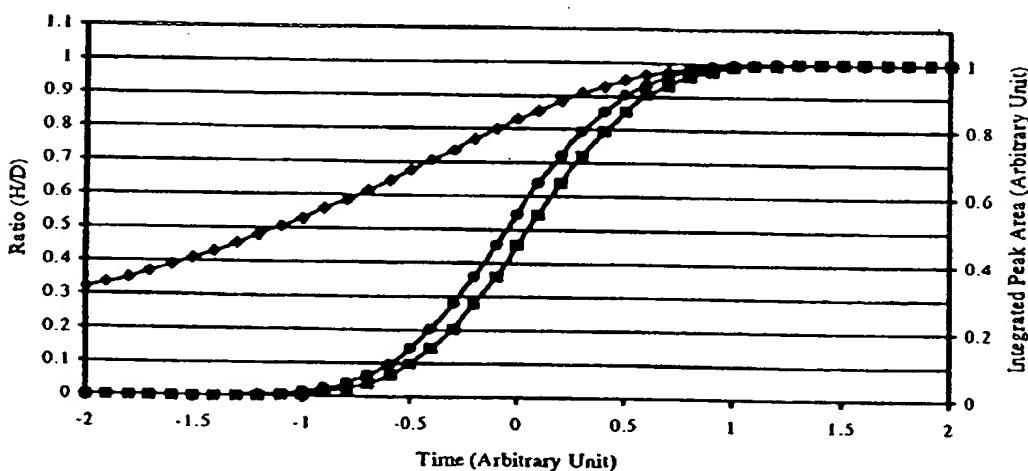
Fig. 2
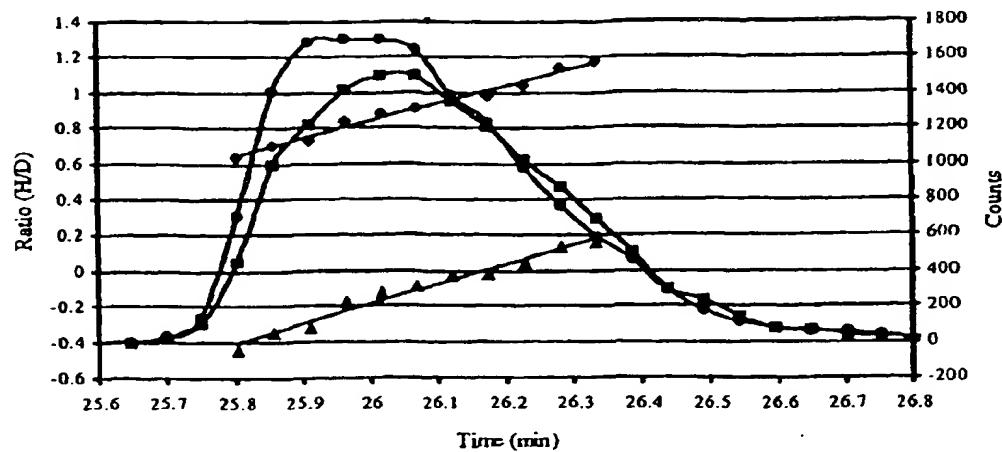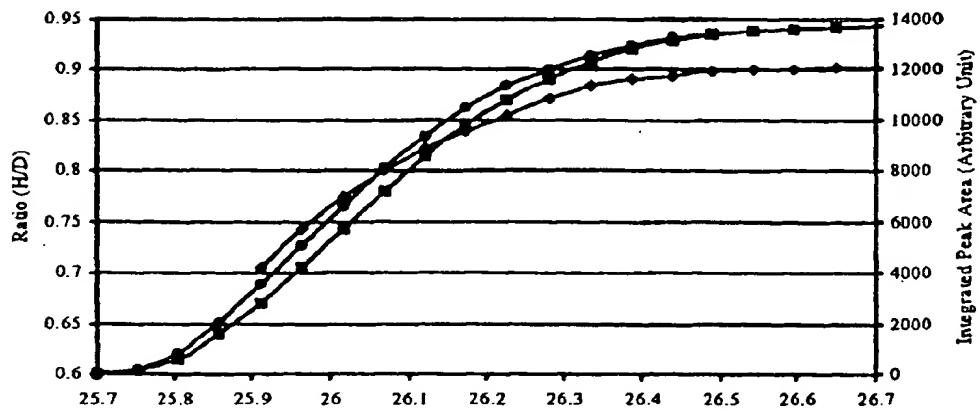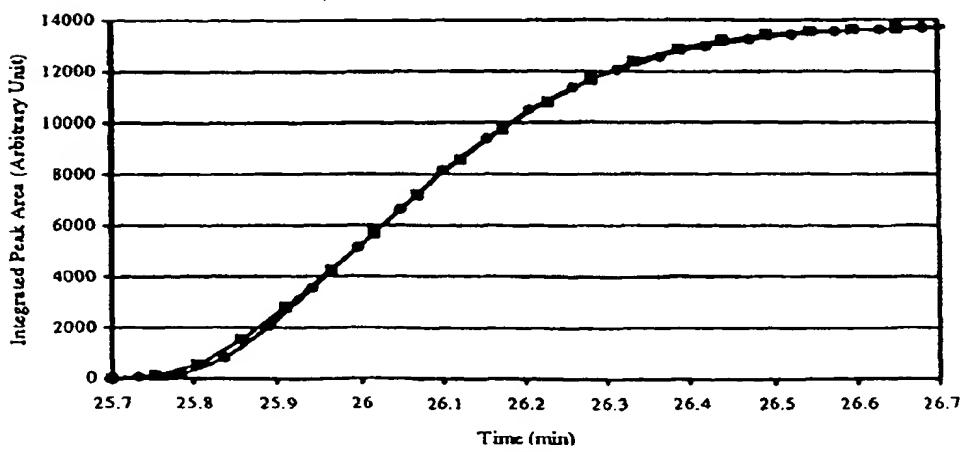
Fig. 3
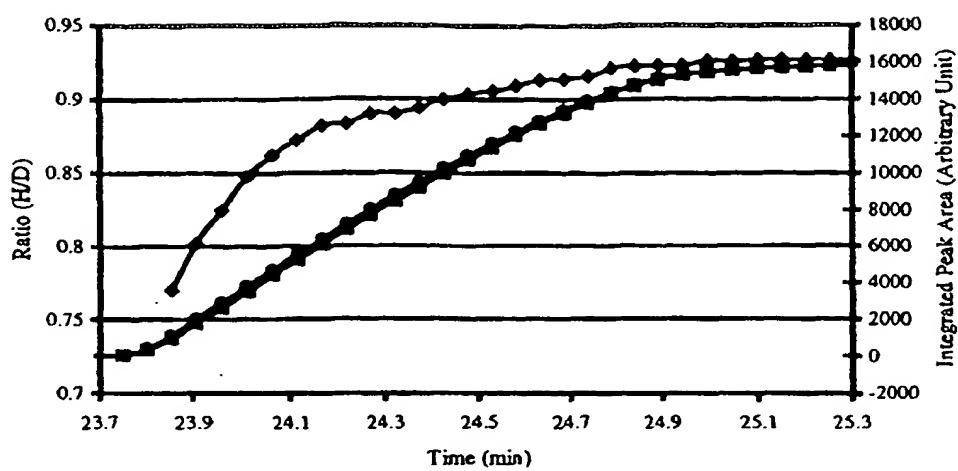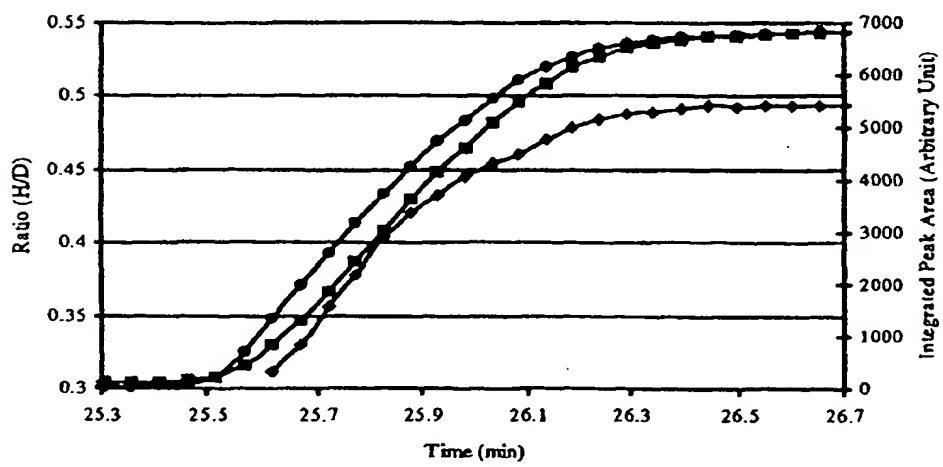
Fig. 4
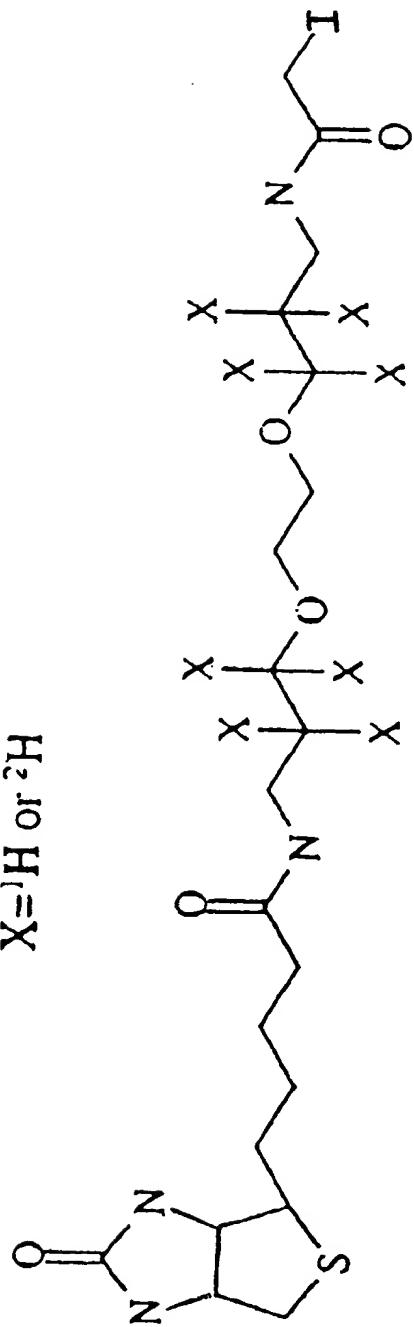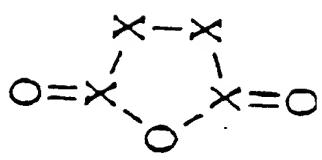
Fig. 5

A
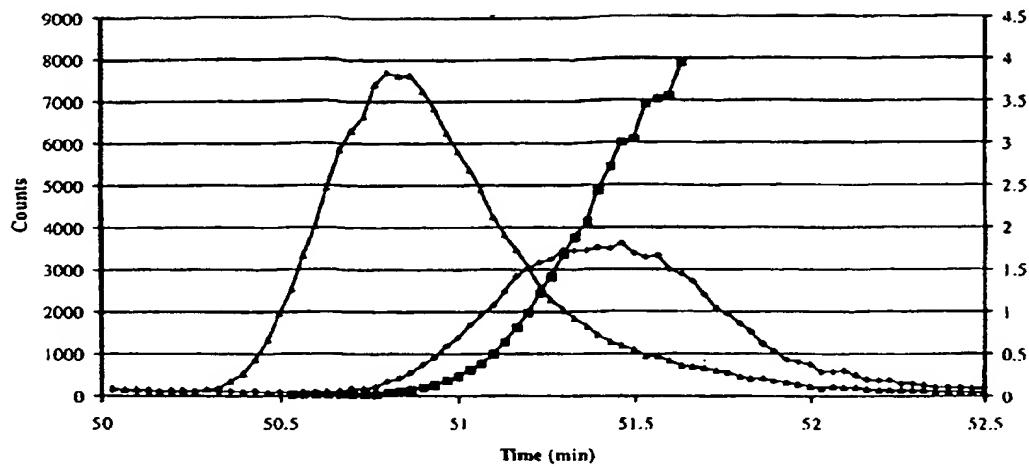
B
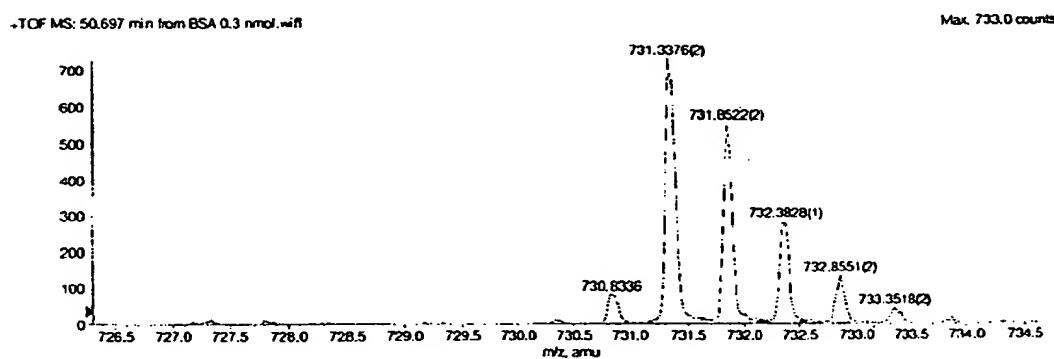
Fig. 6

+TOF MS: 51.197 min from BSA 0.3 nmol.wiff   Max. 830.0 counts
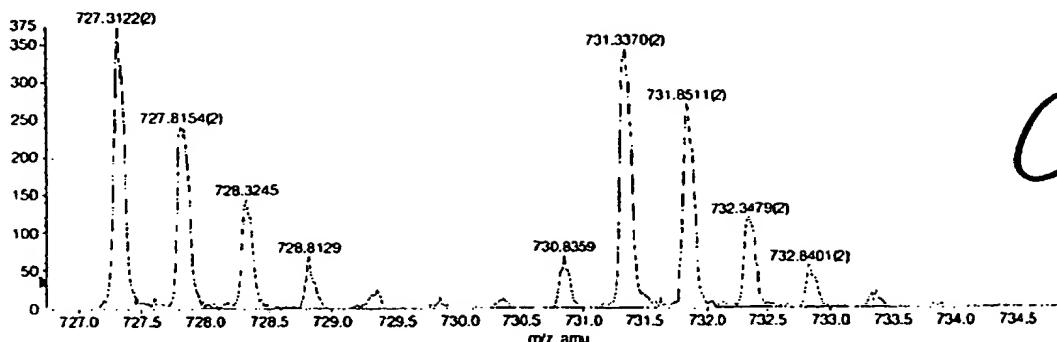
C

+TOF MS: 51.630 min from BSA 0.3 nmol.wiff   Max. 704.0 counts

D

Fig. 6 (cont.)

Fig. 6 (cont.)

A

B
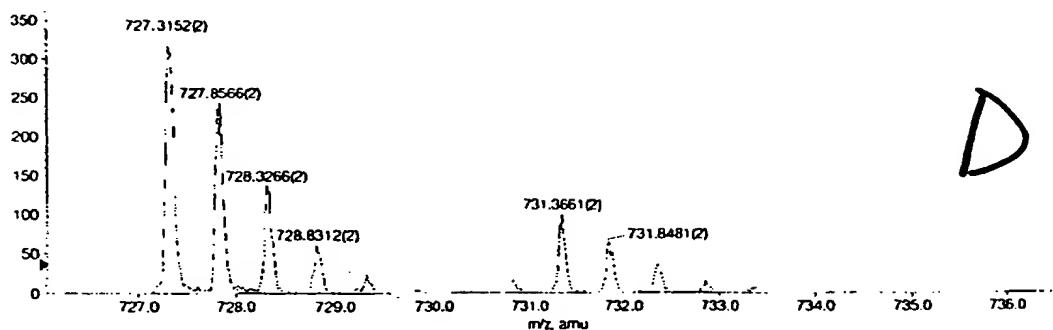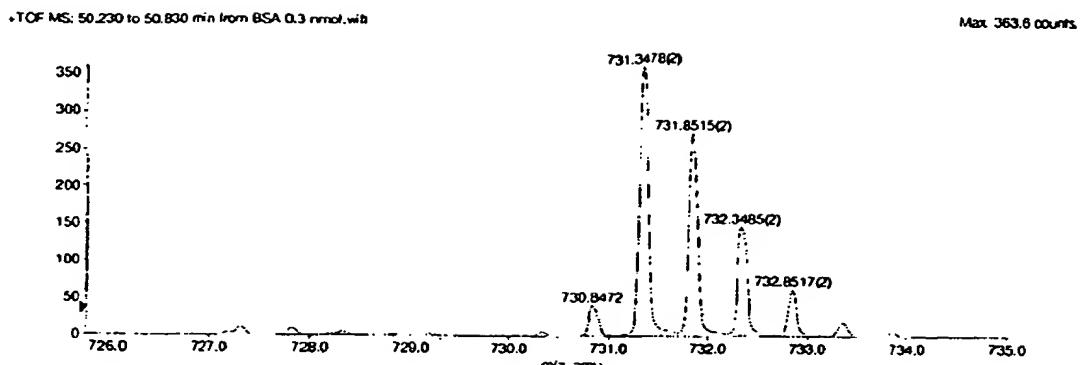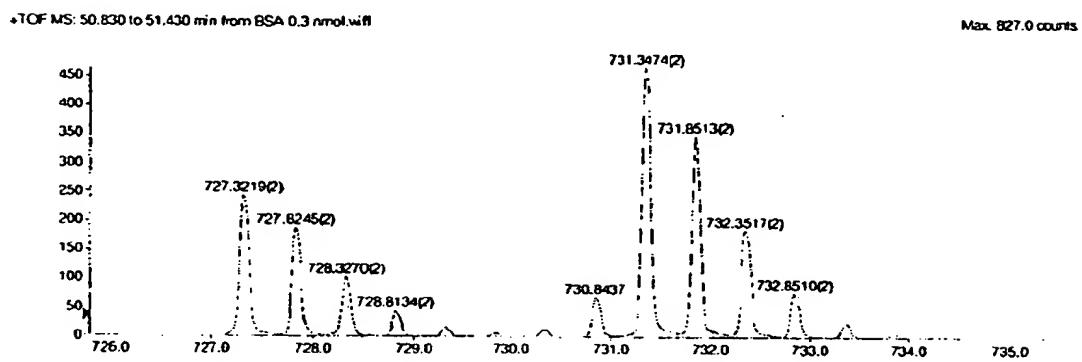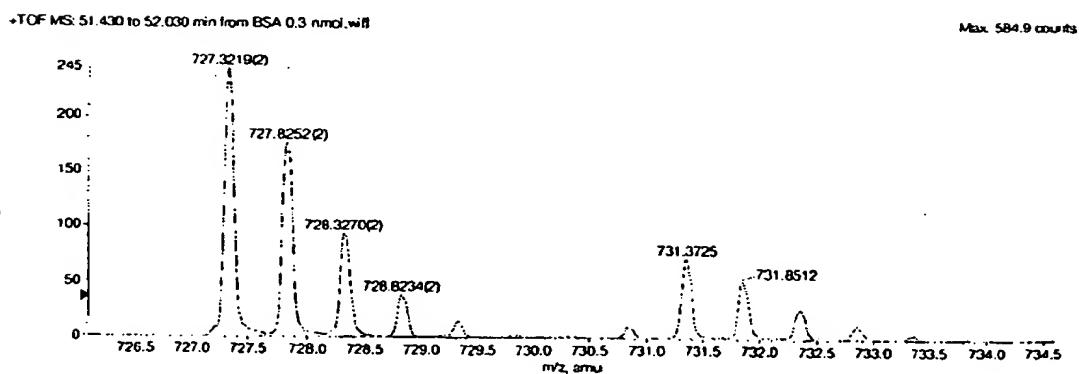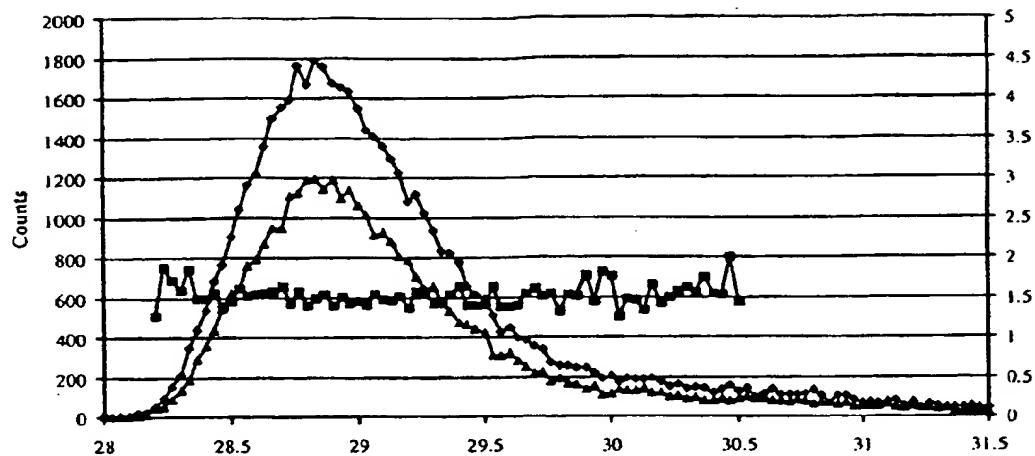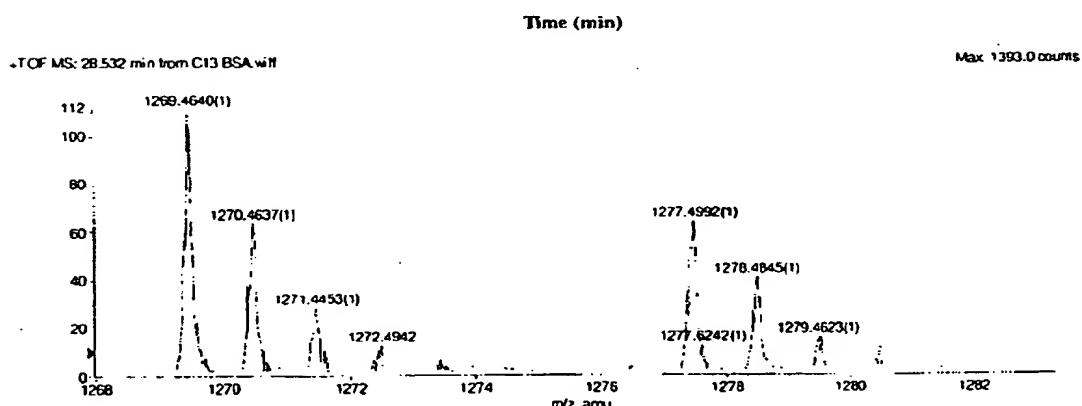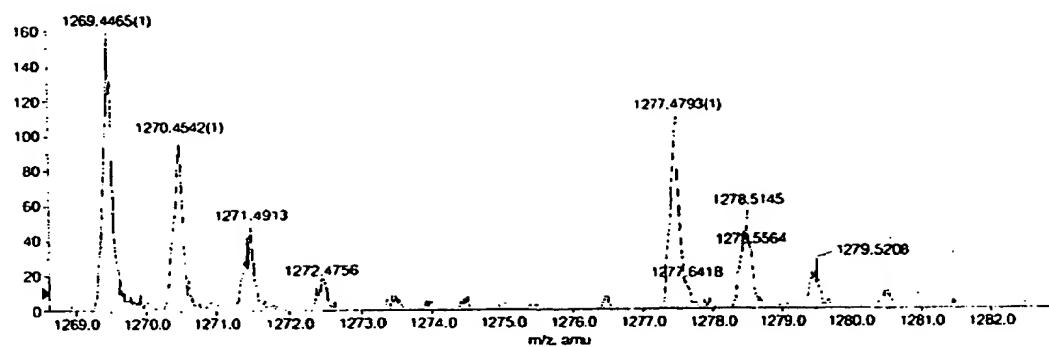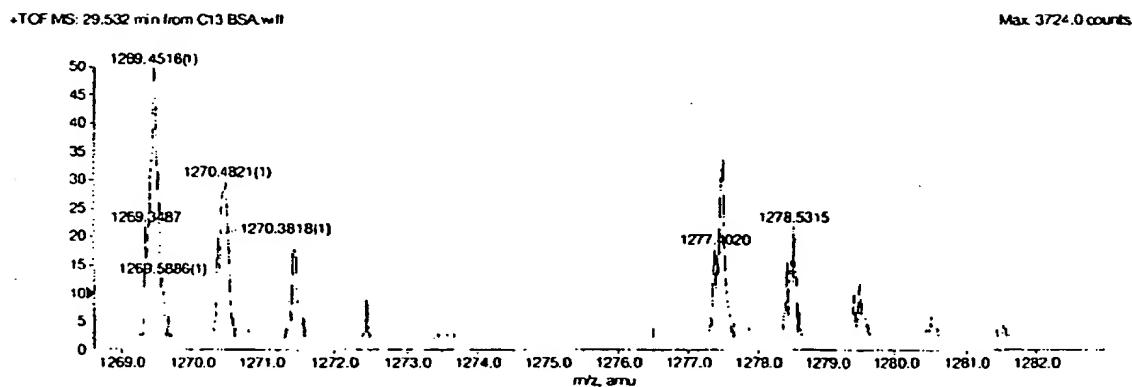
Fig. 7

Fig. 7 (cont.)
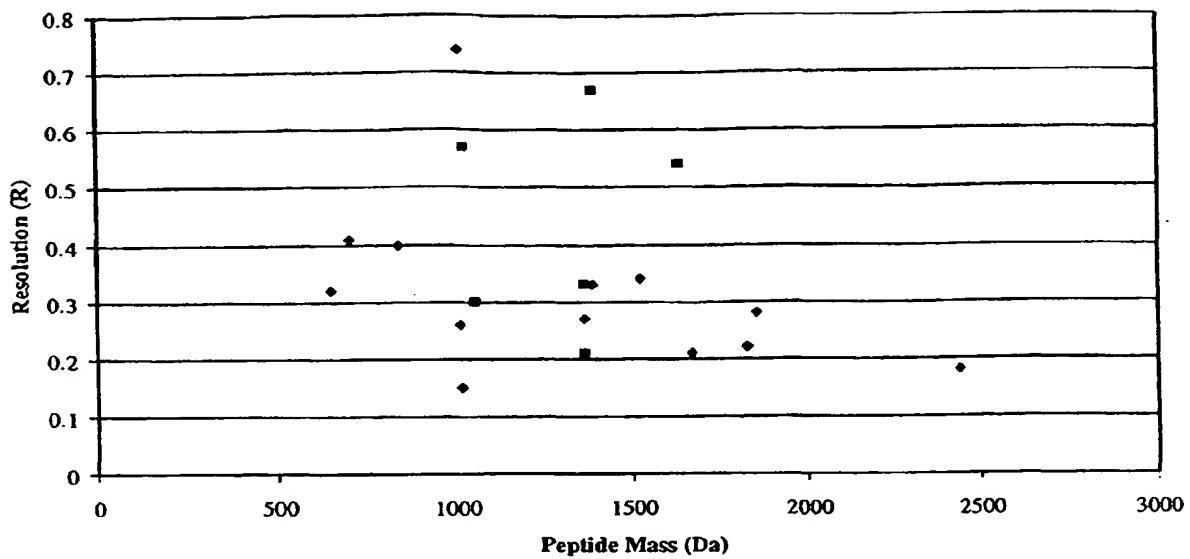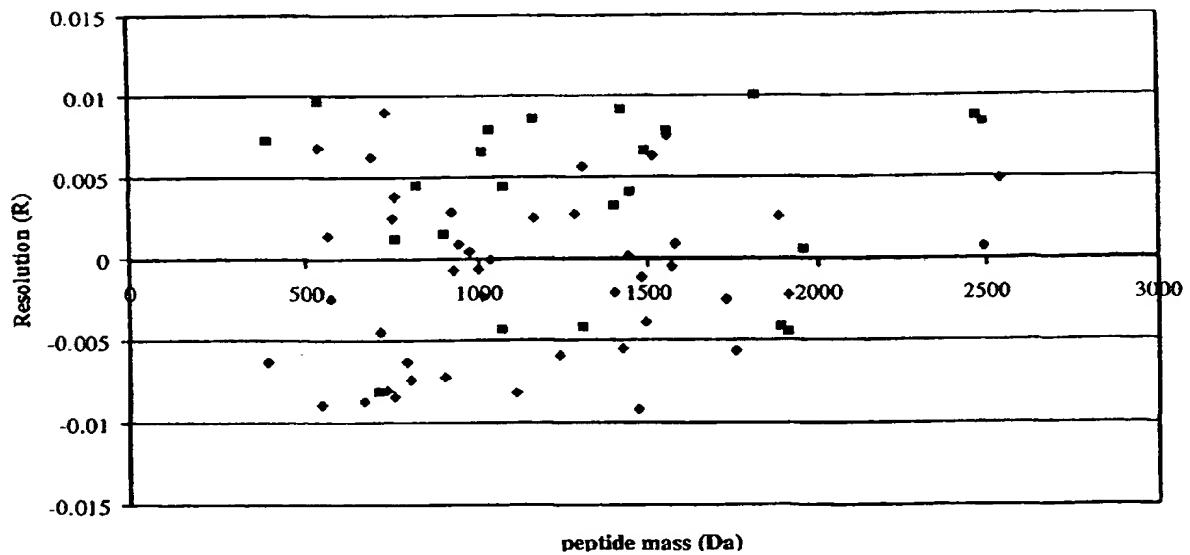
Fig. 8
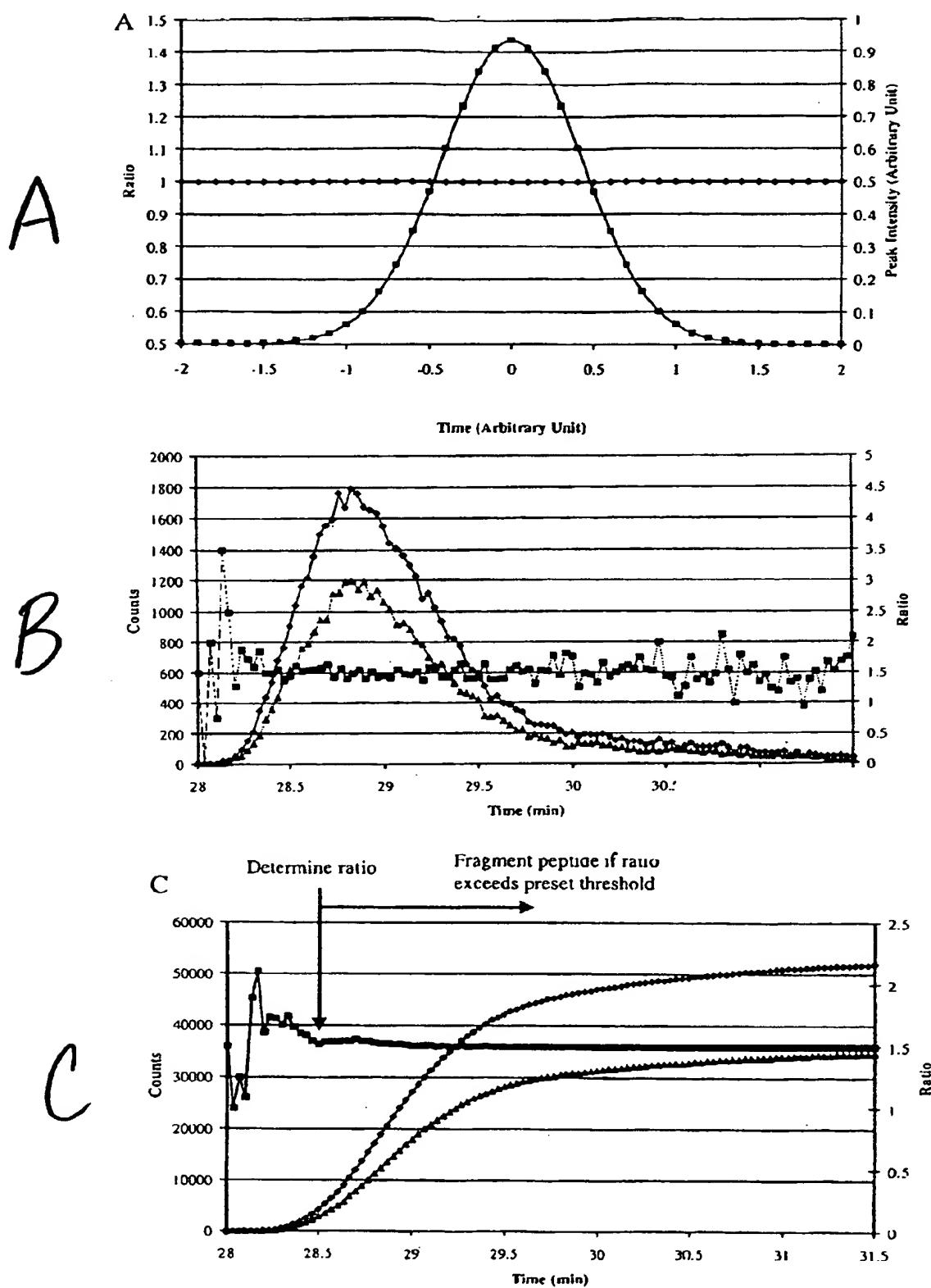
Fig. 9

+TOF MS: 39.025 min from succinic anhydride multiplex.wiff   Max 3002.0 counts.

Fig. 10

+TOF MS: 39.125 min from succinic anhydride multiplex.wiff   Max. 2823.0 counts.

Fig. 10 (cont.)

+TOF MS: 39.359 min from succinic anhydride multiplex.wiff   Max. 1620.0 counts.

Fig. 10 (cont.)

Fig. 11

Fig. 12

C

D

Fig. 12 (cont.)

Fig 12 (cont.)

A

B

Fig. 13

Fig. 14

C

D

E

Fig. 13 (cont.)

(51) International Patent Classification[7]: G01N 33/68, 33/58

(21) International Application Number:
PCT/US2002/030742

(22) International Filing Date:
27 September 2002 (27.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/325,335   27 September 2001 (27.09.2001)   US

(71) Applicant: PURDUE RESEARCH FOUNDATION [US/US]; 1291 Cumberland Avenue, West Lafayette, IN 47906 (US).

(72) Inventors: REGNIER, Fred, E.; 1219 Tuckahoe Lane, West Lafayette, IN 49706 (US). ZHANG, Roujian,; 587 N. Ventu Park Road, Ste E PMB 126, Newberry Park, CA 91320 (US).

(74) Agent: SANDBERG, Victoria, A.; Mueting, Raasch & Gebhardt, P.A., P.O. Box 581415, Minneapolis, MN 55458-1415 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

(88) Date of publication of the international search report:
26 February 2004

[Continued on next page]

(54) Title: CONTROLLING ISOTOPE EFFECTS DURING FRACTIONATION OF ANALYTES

(57) Abstract: Compositions and methods for controlling or eliminating isotope effects during fractionation of chemically equivalent but isotopically distinct compounds. Isotope coding agents contain heavy isotopes other than deuterium. The invention facilitates intelligent data acquisition. After sample fractionation, isotope abundance ratios are calculated using mass spectrometry, and analytes of interest are identified in real time.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   G01N33/68   G01N33/58

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 11208 A (UNIV WASHINGTON) 2 March 2000 (2000-03-02) the whole document	1-63
A	ZHANG R ET AL: "Fractionation of isotopically labeled peptides in quantitative proteomics." ANALYTICAL CHEMISTRY. UNITED STATES 1 NOV 2001, vol. 73, no. 21, 1 November 2001 (2001-11-01), pages 5142-5149, XP002260710 ISSN: 0003-2700 cited in the application the whole document	1-63

—/—

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 November 2003	27/11/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340-2040, Tx. 31 651 epo nl, Fax: (+31–70) 340-3016	Schwachtgen, J-L

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 02 46770 A (GENETICS INST LLC) 13 June 2002 (2002-06-13) the whole document	1-63
P,X	ZHANG ROUJIAN ET AL: "Minimizing resolution of isotopically coded peptides in comparative proteomics." JOURNAL OF PROTEOME RESEARCH. UNITED STATES 2002 MAR-APR, vol. 1, no. 2, March 2002 (2002-03), pages 139-147, XP002260715 ISSN: 1535-3893 the whole document	1-63

INTERNATIONAL SEARCH REPORT

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0011208	A	02-03-2000	AU	755334 B2	12-12-2002
			AU	5691399 A	14-03-2000
			EP	1329513 A1	23-07-2003
			EP	1105517 A1	13-06-2001
			JP	3345401 B2	18-11-2002
			JP	2002523058 T	30-07-2002
			JP	2003107066 A	09-04-2003
			WO	0011208 A1	02-03-2000
			US	2002076739 A1	20-06-2002
WO 0246770	A	13-06-2002	AU	4173202 A	18-06-2002
			CA	2426728 A1	13-06-2002
			WO	0246770 A2	13-06-2002
			US	2003054570 A1	20-03-2003